이학박사 학위논문

Genomic and transcriptomic analysis of 180 well differentiated thyroid neoplasms and 16 anaplastic thyroid carcinomas using massively parallel sequencing

차세대 염기서열 분석을 활용한 분화갑상선종양 180 례와 미분화갑상선암 16 례의 유전체 및 전사체 분석 연구

2018 년 2 월

서울대학교 자연과학대학원

협동과정 생물정보학

유 승 근

Genomic and transcriptomic analysis of 180 well differentiated thyroid neoplasms and 16 anaplastic thyroid carcinomas using massively parallel sequencing

By

Seong-Keun Yoo

February 2018

Interdisciplinary Program in Bioinformatics

Seoul National University

차세대 염기서열 분석을 활용한
분화갑상선종양 180례와
미분화갑상선암 16례의
유전체 및 전사체 분석 연구

지도교수 천 종 식

이 논문을 이학박사 학위논문으로 제출함
2018년 2월

서울대학교 자연과학대학원
협동과정 생물정보학전공
유 승 근

유승근의 이학박사 학위논문을 인준함
2017년 12월

위 원 장 _____ (인)
부위원장 _____ (인)
위 원 _____ (인)
위 원 _____ (인)
위 원 _____ (인)

ABSTRACT

Genomic and transcriptomic analysis of 180 well differentiated thyroid neoplasms and 16 anaplastic thyroid carcinomas using massively parallel sequencing

Seong-Keun Yoo

Major in Interdisciplinary Program in Bioinformatics

College of Natural Science

Seoul National University

Well-differentiated thyroid carcinoma (WDTC) is the most prevalent human endocrine cancer with good prognosis. Recently, The Cancer Genome Atlas (TCGA) broadened the understanding of this disease by analyzing data for 496 papillary thyroid carcinomas (PTCs) which is the most common type of WDTC. However, in their analysis, follicular thyroid carcinoma (FTC) which accounts for 10% of thyroid malignancy was not included. Furthermore, the molecular pathogenesis from WDTC to anaplastic thyroid carcinoma (ATC) which is a very rare and one of the most lethal type of human malignancy is

not fully elucidated until now. In the first chapter, the molecular characteristics of WDTCs were illustrated by large scale RNA-sequencing (RNA-seq) analysis on 180 WDTCs including 30 minimally invasive FTCs (miFTCs), 25 follicular adenomas (FAs), 48 follicular variant of PTCs (FVPTCs), and 77 classical PTCs (cPTCs). The mutation profiling suggested that the gene alterations in WDTC is very mutually exclusive and the type of driver mutation was tightly associated with the histology of thyroid cancer. *BRAF*^{V600E} was only identified in PTC (71.43% and 25.00% for cPTC and FVPTC, respectively) and most fusion genes were discovered in PTC (17.60%) rather than miFTC or FA (only one *PAX8-PPARG* in miFTC). Meanwhile, *H/K/NRAS* mutations were found in FVPTC, miFTC, and FA with high incidence (47.92%, 50.00%, and 24.00%, respectively). Moreover, the alterations in *DICER1*, *EIF1AX*, *EZH1*, *IDH1*, *PTEN*, *SOS1*, and *SPOP* were frequently identified in miFTC and FA rather than PTC. Based on gene expression profiles, WDTCs were classified as three molecular subtypes, Non-*BRAF*-Non-*RAS* (NBNR) as well as *BRAF*-like and *RAS*-like regardless of their histological classification. Each molecular subtype represented the differential chromosomal aberration and regulation of signaling pathways such as MAPK, p53, cell-cell communication, and metabolism pathways. In the second chapter, to describe genomic, transcriptomic, and epigenomic features of ATC, whole-genome sequencing with *HpaII* digestion and RNA-seq were performed on 16 ATCs. Moreover, the integrative analysis with additional 324 WDTCs demonstrated the genomic, transcriptomic, and epigenomic changes during the progression of ATC from WDTC. *TERT*

(75%), *BRAF* (43.75%) *H*/*NRAS* (31.25%) were the most recurrently altered genes in ATC. In addition, 75% of ATC had extra somatic mutation or deletion in *EIF1AX*, *AKT1*, *PIK3CA*, *TP53*, *CDKN2A*/*CDKN2B*, *BRCA1*, *PTEN*, *TET1*, *LATS1*/*LATS2, ATRX*, *KMT2C*, and *U2AF1*. The patients with ATC harbored *TP53* germline mutations more frequently than WDTC patients (p = 0.002). Through gene expression analysis, ATCs were separated from WDTCs as a novel molecular subtype, ATC-like. Notably, the significant up-regulation of *IL6*, *CD274* (PD-L1) and *PDCD1LG2* (PD-L2) were observed which could be considered as potential immunotherapeutic targets. Interestingly, *CD274* and *PDCD1LG2* were usually up-regulated in ATCs with deep deletion of *CDKN2A*/*CDKN2B* and they represented the worst thyroid cell differentiation. From DNA methylation analysis, the global DNA methylation pattern of ATC was distinguished from WDTC and the more frequent hypermethylation of ATC, enriched in CpG island (p < 0.001), than WDTC was detected. These significant results broadened the current understanding of thyroid carcinoma hence allows more efficient diagnostic strategy and attractive targeted therapeutic targets in the future.

* The first chapter was published in Plos Genetics [1].

Keywords: Thyroid carcinoma; Well-differentiated thyroid carcinoma; Follicular thyroid carcinoma; Papillary thyroid carcinoma; Anaplastic thyroid carcinoma; Massively-parallel sequencing

Student number: 2014-21325

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

MPS: Massively-parallel sequencing

bp: Base pair

WGS: Whole-genome sequencing

WES: Whole-exome sequencing

RNA-seq: RNA sequencing

Gb: Gigabase

SNV: Single nucleotide variant

Indel: Short insertion/deletion

CNA: Copy number alteration

SV: Structural variation

TSG: Tumor suppressor gene

DEG: Differentially expressed gene

TCGA: The cancer genome atlas

Mb: Megabase

SCNA: Somatic copy number alteration

RD: Read depth

PCA: Principal component analysis

WDTC: Well-differentiated thyroid carcinoma

PTC: Papillary thyroid carcinoma

FTC: Follicular thyroid carcinoma

MTC: Medullary thyroid carcinoma

FA: Follicular adenoma

ATC: Anaplastic thyroid carcinoma

PDTC: Poorly-differentiated thyroid carcinoma

FC: Fold change

FVPTC: Follicular variant of papillary thyroid carcinoma

FTN: Follicular thyroid neoplasm

miFTC: Minimally invasive follicular thyroid carcinoma

cPTC: Classical papillary thyroid carcinoma

PCR: Polymerase chain reaction

GATK: Genome analysis toolkit

BQSR: Base quality score recalibration

ExAC: Exome aggregation consortium

MAF: Minor allele frequency

rlog: Regularized log

TDS: Thyroid differentiation score

JRB: Jointly-regulated block

FPKM: Fragment per kilobase per million

ANCOVA: Analysis of covariance

LT: Lymphocytic thyroiditis

NBNR: Non-*BRAF*-Non-*RAS*

LNM: Lymph node metastasis

ETE: Extrathyroidal extension

EFVPTC: Encapsulated follicular variant of papillary thyroid carcinoma

TCA cycle: Citric acid cycle

KEGG: Kyoto encyclopedia of genes and genomes

CAM: Cell adhesion molecule

ECM: Extracellular matrix

OXPHOS: Oxidative phosphorylation

Chr: Chromosome

CDS: Coding DNA sequence

CTX: Inter-chromosomal translocation

UTR: Untranslated region

wiFTC: Widely invasive follicular thyroid carcinoma

WGCNA: Weighted gene co-expression analysis

VST: Variance stabilizing transformation

DSS: Disease-specific survival

PT: Primary tumor

LRT: Locally recurred or residual tumor

DMT: Distant metastatic tumor

LN: Lymph node

DOD: Death of disease

AWD: Alive with disease

FU: Follow-up

NA: Non-applicable

CCLE: Cancer cell line encyclopedia

RPKM: Read per kilobase per million

General Introduction

Massively-parallel sequencing

Massively parallel sequencing (MPS) or next-generation sequencing is adopted as a gold standard method for understanding human genome in recent 10 years [2]. It is considered as a second generation sequencing after Sanger and colleagues developed a chain termination sequencing (also known as Sanger sequencing) in 1977 [3]. The unique characteristics of this technique are as follows: i) the preparation of sequencing libraries without bacterial cloning of DNA fragment, ii) thousands of reactions are arose in parallel, and iii) unnecessity of electrophoresis for detect sequencing result [2]. The aforementioned advantages of MPS technology enables the generation of massive number of sequence data which covers individual genome with up to hundreds of times.

The most widely adopted platform of MPS is Illumina sequencing technology which implements bridge amplification approach [4]. It results in up to 250 base pair (bp) length of paired-end reads. As it has short read length, it is often referred to as a short read sequencing. Recently, the third generation sequencing platforms which generate long reads length up to 5 kilobases were introduced in this field [5, 6]. However, their huge sequencing cost and processing time are challenges for researchers. Thus, MPS is still the best choice for sequence analysis with a large number of sample. There are various types of MPS methods such as whole-genome sequencing (WGS), whole-exome sequencing (WES), and RNA sequencing (RNA-seq). As each method

has its pros and cons, the integration of various method for appropriate analysis provides more comprehensive understanding of molecular features of an organism.

WGS generates sequenced bases up to hundreds of gigabase (Gb) level. This method facilitates investigators to accurately discover sequence variation including single nucleotide variant (SNV), short insertion/deletion (indel), copy number alteration (CNA), and structural variation (SV) [7, 8]. As it covers the entire genome with uniform coverage, the most advantageous part of this method is identification of the precise breakpoint of CNA and SV. However, its expensive cost and data size are huge obstacles for investigators. Thus, WES which covers only protein coding regions (exome) was developed and allowed more effective study designs [9].

Although the association between regulatory sequences and human diseases is increasingly analyzed [10-12], the majority of variation which contributes to phenotypic changes is discovered in exome. Since WES only generates sequencing data within exome which accounts for approximately 1% of genome, the deeper sequencing coverage than WGS, over hundreds, could be accomplished with cheaper price. Therefore, WES allows the most accurate and effective identification of SNV and indel with clinical implications [13-15].

RNA-seq is a revolutionary tool for transcriptome profiling [16]. It enables researchers to quantify gene expression and identify alternative splicing. Furthermore, the sequence variation including SNV and indel in transcribed mRNA could be discovered. This tool also allows identification of novel transcript such as unannotated protein coding genes or lncRNA [17, 18], and it could be accomplished by *de novo* assembly of whole-transcriptome data [19, 20]. The most attractive usage of this method is comparing the expression level of genes or transcript between groups under different conditions [21-23].

Due to the above mentioned advantages and dramatic cost reduction, MPS is widely adopted in diverse clinical research fields and/or applications including the molecular diagnosis of rare disease, non-invasive prenatal diagnosis, pre-implantation genetic diagnosis, and even cancer genomics [24-27].

Cancer genomics

Cancer is a disease of the genome [28]. It arises by the somatic mutations in oncogene or tumor suppressor gene (TSG). In the past, linkage analysis has been the prior method to identify cancer susceptibility genes such as *BRCA1* or *BRCA2* [29, 30]. Lately, with privilege of the development of MPS technology, researchers could extensively discover not only somatic mutations, but also the transcriptomic and epigenetic alterations in cancer genome. Particularly, WGS and WES enabled researchers to identify actionable mutations in oncogene of cancer genome [27, 31]. Furthermore, RNA-seq is the most powerful technology in cancer genomics study which could discover molecular subtypes, differentially expressed genes (DEGs) between those subtypes, and oncogenic fusion gene drivers [32-34].

The largest scale of cancer genomics study is The Cancer Genome Atlas (TCGA) project. It launched on 2005 and the studies from TCGA allowed us to bring our knowledge on characteristics of cancer to the next level [35-40]. They integrated MPS (WGS, WES, and RNA-seq) and microarray (single nucleotide polymorphism, methylation, and protein array) methods to understand the underlying mechanism of cancer.

In general, cancer genomics study starts from discovering driver mutations (SNV and indel) in oncogene or TSGs such as *BRAF*, *EGFR*, *H/K/NRAS*,

TP53, and *PTEN*. There are varying numbers of somatic mutations in cancer genome up to hundreds of mutations per megabase (Mb) [41]. However, most of them do not promote the malignant transformation, thus often referred as 'passenger mutation' [42]. By the way, only one or few of somatic mutations in oncogenes (e.g. *BRAF* and *H/K/NRAS*) confer growth advantage and drive the malignancy of cells, hence it is called as 'driver mutation' [42]. Mostly, the actionable (usually in oncogene) or truncating (usually in TSG) mutations are located in protein coding regions, exome [43]. Therefore, WES is considered as the prior method since it produces the most uniform and highest sequencing throughput across exome with low cost relative to WGS and RNA-seq.

To identify somatic driver mutations, the most efficient approach is sequencing patient's tumor and its adjacent normal tissue or blood samples. The software such as MuTect judge the variant as somatic when the variant found in tumor sample is not present in adjacent normal tissue [44]. However, when the somatic mutations are present in data from adjacent normal sample due to tumor contamination, they could be filtered out from variant calls. Thus, the careful data interpretation is required. The accurate somatic mutation identification is significant since each tumor reacts differentially to anti-cancer drugs depending on its driver or passenger mutation although they share similar histological characteristics [45-47].

Recently, there have been attentions to passenger mutations in cancer genome as a rich source of historical information [48]. Although these mutations do not contribute to cancer development, their unique combinations indicate the different source of mutational processes [41]. There are several exogenous and endogenous mutational processes which generate somatic mutations such as tobacco smoking, ultraviolet, and DNA replication deficiency [41]. The pattern of somatic mutations in caner genome is referred to as a 'mutational signature' [41]. It could be briefly measured with six nucleotide substitution subtypes: C>A, C>G, C>T, T>A, T>C, and T>G [41]. More recently, the expanded concept of the mutational signature was introduced and it covers 96 types of mutation context such as C>A at NpCpN trinucleotides [41]. Recent reports also described the clinical implications of the mutational signature in certain cancers as well as the association between mutational process [49, 50]. Therefore, the mutational signature analysis is one of the most significant field of cancer genomics.

In addition to somatic SNV and indel, oncogenic fusion gene rearrangement is another important event which drives malignant transformation [33]. This event arises by inter-chromosomal translocation or CNAs such as large deletion and duplication of genome [51-53]. The oncogenic fusion gene rearrangements frequently incorporate kinase gene fusion such as *ALK*, *BRAF*, *RET*, *FGFR2*, *NTRK1*, and *NTRK3* [54]. In addition to small size mutations, they are also considered as important anti-cancer therapeutic targets [55]. Aforementioned genes are fused with several partner genes across genome-

wide level. With the advancement of MPS technology, especially RNA-seq, the novel oncogenic fusion gene rearrangements and partner genes are actively discovered since this method could sequence adequately expressed mRNA across genome [33, 54]. However, the actual breakpoint of fusion gene rearrangement could not be predicted by RNA-seq because it does not cover intronic regions where the most of breakpoints are located. To overcome this problem, WGS is also being used for fusion gene rearrangement discovery with accurate breakpoint detection [56].

The somatic CNA (SCNA) is another important genetic event which drives tumorigenesis and contributes the aggressiveness of cancer [57, 58]. The two types of SCNA are being examined widely in recent studies, focal and arm-level SCNA [57]. Focal SCNA usually denotes the deletion or amplifications of regions with particular genes such as TSG or oncogene, respectively [57]. Furthermore, the arm-level SCNA represents the aneuploidy of the chromosome arms and could be applied as a prognostic indicator [59, 60]. The discovery of SCNA also required tumor and adjacent normal tissue samples. There are several SCNA detection methods using WES technology, thus WES gives an opportunity to analyze not only somatic SNV and indel, but also SCNA in cost effective manner. The most popular algorithm for CNA detection in WES is a read depth (RD) approach [61]. The basic RD based CNA detection algorithm implements two steps: preprocessing and segmentation [61]. The segmentation steps are generally performed with circular binary algorithm [62]. It generates large chromosome segment when

each captured region has similar copy number values until adjacent genomic regions with different copy number is appeared [61]. The major problem of CNA detection using WES is that the real copy number within non-exonic regions (intronic and intergenic regions) could not be accurately measured by this method, because WES only covers 1% of genome. Therefore, the copy number status of aforementioned regions was only predicted by CBS algorithm and it could produce biased result. Recently, to overcome this issue, there are some efforts that measure copy number values of non-exonic regions using off-target reads [63, 64]. Meanwhile, WGS is the most accurate approach for SCNA detection which do not have aforementioned limitation, because it could generate uniform copy number data across genome.

After discovery of the genomic events from cancer genome, recent studies make a tremendous effort to reveal the molecular subtypes of each cancer type. The molecular classification of cancer is often achieved by transcriptome profiling. The several analytic methods are applied to this step including non-negative matrix factorization, principal component analysis (PCA), and hierarchical clustering [65, 66]. The molecular subtypes of cancer are tightly associated with the clinical significance such as overall survival because each of them represent differential signaling activation or repression [39, 50, 67, 68]. Furthermore, DNA methylation, protein, miRNA, and lncRNA profiles also provide more specified subclasses of cancer [39, 50]. Aforementioned analyses give opportunity to more careful and personalized management of patients depends on their molecular characteristics.

Genomic characteristics of well-differentiated thyroid carcinoma

Thyroid cancer is the most common malignancy in human endocrine system and the incidence rate of female is three times higher than that of male [69, 70]. In recent decades, the incidence rate of this disease is rapidly increasing worldwide [71]. The diagnosis of thyroid cancer is annually increased to 22.7% and 20.6% for male and female between 1999 and 2013, respectively [72]. The remarkable increase of thyroid cancer incidence is often explained by over-diagnosis and over-treatment [73]. In particular, the incidence rates of this disease are obviously elevated in high-income countries such as United States, Republic of Korea, and United Kingdom [73, 74]. The 5-year relative survival rate of thyroid cancer is 100.2% in Korea [72]. Therefore, thyroid cancer is usually considered as a low risk malignancy.

The most common type of thyroid cancer is well-differentiated thyroid carcinoma (WDTC) which includes papillary thyroid carcinoma (PTC) and follicular thyroid carcinoma (FTC) [75]. PTC and FTC account for up to 84% and 10% of all thyroid carcinoma, respectively [76]. These cancers arise from thyroid follicular epithelial cells in contrast to medullary thyroid carcinoma (MTC) which is originated from the parafollicular C cells of the thyroid [77]. Since they maintain the important ability of the normal thyroid cells, such as iodine uptake and accumulation, surgery with radioactive iodine (^{131}I) could be applied without alternative [78, 79]. However, this type of cancer also

displays advancement or metastasis [79]. Ten percent of patients with WDTC display distant metastasis, and it is the leading cause of death by WDTC [80]. The most prevalent metastatic sites are lung, bone, and other sites [80]. In contrast to patients with WDTC, ^{131}I therapy does not work well in those with advanced WDTC [79, 80]. Therefore, it is important to discover the molecular markers to predict and monitor the advancement of this disease.

In 2014, TCGA published the first large scale genomic study of thyroid carcinoma; classical, tall cell variant, and follicular variant types of PTC were analyzed [39]. They discovered well-known oncogenic drivers such as *BRAF*, *HRAS*, *KRAS*, *NRAS*, *CCDC6-RET*, *NCOA4-RET*, *PAX8-PPARG*, and *ETV6-NTRK3*. Moreover, *EIF1AX*, *PPM1D*, and *CHEK2* were also proposed as novel driver genes in PTC. The major finding of this study was novel molecular classification of PTC. They implemented *BRAF*^{V600E}*-RAS* score analysis using the most significant DEGs between *BRAF*- and *RAS*-positive samples to differentiated total samples into *BRAF*^{V600E}-like and *RAS*-like. As a result, all PTC samples were classified as two subtype regardless of their histological subtypes.

In 2015, Cost *et al* described the mutational profile of 68 PTC samples [81]. They discovered somatic mutations in well-known driver genes such as *BRAF*, *RAS*, and *RET* rearrangements. Moreover, novel fusion transcript, *WNK1-B4GALNT3*, and somatic mutations in *DICER1*, *MET*, and *VHL* were

identified. Using RNA-seq, two molecular subtypes, *BRAF*-like and *RAS*-like, were validated in this study.

In 2016, Siraj *et al* presented the WES data of 101 PTC samples and the additional 785 samples were examined as validation cohort [82]. Their analysis proposed the impact of alterations in *TG* gene in addition to *BRAF*, *NRAS*, and *HRAS*. According to this study, *TG* was the third most prevalent altered gene (3.05%) and *TG*-mutated patients displayed the significantly poor survival relative to individuals without *TG* mutation. Also, this type of mutation was more frequently found in advanced PTC (12.7%).

Aforementioned studies well described the genomic and transcriptomic characteristics of PTC. However, their analysis excluded FTC which accounts for 10% of thyroid carcinoma as the second most prevalent. Recently, Jung *et al* described the first mutational landscape of FTC as well as a follicular adenoma (FA) [83]. Follicular adenoma is benign lesion which is indistinguishable from FTC based on cytologic, sonographic, or clinical feature [84]. They analyzed the somatic mutations and SCNA event of 14 FA and 13 FTC samples using WES and comparative genomic hybridization microarray, respectively. FA and FTC harbored somatic mutations in *NRAS*, *HRAS*, *EIF1AX*, *EZH1*, and *SPOP*. Their analysis suggested that the evolutionary age of FA is as old as FTC which denotes the development of FTC from FA is stable at genomic level. They also performed RNA-seq, but

did not conducted transcriptome profiling, this method was just implemented to identify fusion gene.

Put together, the molecular characteristics of WDTC is fairly well illustrated by various research groups. However, most analysis was confined to PTC and transcriptome landscape of FTC is not elucidated until now. Therefore, the comparison of molecular features between PTC and FTC has not yet been performed.

Genomic characteristics of anaplastic thyroid carcinoma

Anaplastic thyroid carcinoma (ATC) or undifferentiated thyroid carcinoma is one of the most aggressive solid tumor in human. ATC is a very rare and only accounts for less than 2% of thyroid carcinoma [85]. It represents up to 50% of deaths by thyroid carcinoma and the median survival rate of patient with this aggressive malignancy is approximately four months [86-88]. It is usually developed from preceding lesion of WDTC or poorly differentiated thyroid carcinoma (PDTC), but also arise *de novo* [89, 90]. It is known that tall cell variant of PTC, which is the aggressive one, is most common to coexist with ATC and conventional or oncocytic FTC are followed [84, 91, 92].

ATC do not display the biological characteristics of normal follicular thyroid cells including iodine uptake, and thyroglobulin synthesis [89]. The patients with ATC present with a rapidly enlarging anterior neck mass, dyspnea, dysphagia, and voice change [93, 94]. At the time of diagnosis, patients often display aforementioned characteristics with systemic symptoms such as anorexia, shortness of breath, and weight loss [95]. Furthermore, the majority of patient with ATC harbored distant metastasis and it is the main cause of death [96, 97]. The most common metastasis sites are lung, bone, and brain [98]. Because of its aggressive features, early diagnosis of this lethal malignancy is important for the patient survival. However, the majority of patients with ATC are diagnosed at advanced stage or have distant metastasis

to other organs which leads to poor prognosis with resistant to conventional therapies in the present [89].

However, beside the clinical implication of ATC, there is insufficient number of large scale study regarding ATC with MPS technology. Kunstman *et al* reported the first mutational landscape of 22 ATC tumors and four ATC cell lines using WES [99]. They identified that *BRAF* and *RAS* mutations were mutually exclusive in ATC and the cell lines with those mutations harbored co-mutation in *TP53*. Moreover, the mutations in *MTOR*, *NF1*, *NF2*, *MLH1*, *MLH3*, *MSH5*, *MSH6*, *ERBB2*, and *USH2A* which are not previously reported in thyroid cancers were discovered.

Kasaian *et al* analyzed one primary ATC tumor and three ATC cell lines using MPS technologies [100]. They discovered *TP53* and *BRAF* as the most prevalent mutations and novel fusion genes including *MKRN1-BRAF*, *FGFR2-OGDH*, and *SS18-SLC5A11*. Furthermore, they found that the aneuploidy and large size copy number gain and loss are distributed across whole-genome. They also compared the transcriptome of four samples to 58 PTC and matched normal thyroid samples from TCGA study. Based on this analysis, they reported up-regulation of mTOR signaling pathway and down-regulation of druggable targets including *FGFR*s, *VEGFR*s, *KIT*, and *RET*. It was the first study with transcriptome profiling using RNA-seq, but the power of analysis was insufficient, because there is only one primary ATC sample.

More recently, Landa *et al* reported a study regarding genomic and transcriptomic hallmarks of PDTC and ATC [101]. They found recurrent *TERT* mutations in promoter region in addition to *RAS*/*EIF1AX* co-mutation and various cancer-related genes such as *AKT1*, *PIK3CA*, *STK11*, and *NF1*. Moreover, their gene expression analysis revealed that *RAS*-positive ATC represents similar differentiation and MAPK activation as *BRAF*-like ATC. Furthermore, they illustrated high infiltration of M2 macrophage in ATC using 68 gene signatures. This study is the most well-established genomic features of PDTC and ATC, but their targeted sequencing approach restricted to discover only mutations in well-known cancer-related genes. Moreover, microarray based transcriptome analysis have limitation in resolution; low abundance transcript and much higher fold change (FC) could be more accurately detected by RNA-seq [102]. In addition to aforementioned studies, there were three studies which covered small number of genes using targeted sequencing method [103-105].

Put together, the genomic characteristics, especially the mutational landscape, of ATC is well investigated by recent efforts. However, their analysis mainly covered the limited number of genes and did not include large scale transcriptome analysis with high resolution method such as RNA-seq. Moreover, most of them did not fully characterized focal SCNA in ATC; the

majority of analysis was restricted to the arm-level SCNA [99, 101]. Therefore, the molecular characteristics of ATC is not fully revealed yet.

Objectives of this study

There have been incredible efforts in deciphering the molecular characteristics of WDTC but recent reports were generally focused on PTC [39, 81, 82]. Therefore, the comprehensive analysis on WDTC is not completely achieved because there is not adequate number of research regarding FTC which displays more frequent hematogeneous spread and second prevalent form [84]. Also, benign FA which is hard to distinguished from FTC before surgery is not also well investigated in the molecular level [84]. Moreover, the integrated analysis of ATC with WDTC to illustrate the molecular pathogenesis of this lethal disease has not been performed yet. Recent studies well described the genomic hallmarks of ATC, but their analysis was restricted to small region of genome and transcriptome landscape was not fully investigated [99-101, 103-105]. The integrative analysis on aforementioned type of thyroid cancer, PTC, FTC, and ATC, as well as benign FA, could provide more clinically significant interventions about early diagnosis and preventing the progression of these malignancies. Thus, it would lead to the reduction of over-diagnosis and over-treatment, in addition to prolong the patient survival.

In this thesis, various MPS technologies including WGS, WES, and RNA-seq were applied to illustrate the overall landscape of thyroid carcinoma from FA which is a benign lesion to ATC which is the most aggressive human cancer. The newly sequenced and publicly available data from hundreds of thyroid

carcinoma samples were analyzed in-depth approach. WGS, WES, and RNA-seq were implemented to identify the somatic mutations, SCNAs, mutational signatures, and molecular subtypes. It is expected that this study will allow us to have more comprehensive understanding of molecular pathogenesis of this disease and would lead to discover novel molecular diagnostic and therapeutic targets.

CHAPTER 1.

Comprehensive analysis of the transcriptional and mutational landscape of 180 follicular and papillary thyroid neoplasms

Abstract

The well-differentiated thyroid carcinoma (WDTC) is mainly composed of papillary thyroid carcinoma (PTC; ~80%) and follicular thyroid carcinoma (FTC; ~10%). In 2014, The Cancer Genome Atlas (TCGA) described the genomic landscape of WDTC by analyzing 496 PTCs. However, their analysis did not include FTC which is the second most common type of WDTC. Therefore, the molecular characteristics of WDTC is not fully revealed yet. Benign follicular adenoma (FA) is indistinguishable from FTC by preoperative diagnosis due to their similar histological features, thus it causes diagnostic inconvenience in clinic. In this thesis, the first large scale RNA sequencing analysis was conducted, with data for 30 minimally invasive FTCs (miFTCs) and 25 FAs, to discover their difference on molecular level. Also, 77 classical PTC (cPTCs) and 48 follicular variant of PTCs (FVPTCs) were sequenced to compare the molecular properties of papillary- and follicular-patterned thyroid tumors. Mutations in *H/K/NRAS*, *DICER1*, *EIF1AX*, *IDH1*, *PTEN*, *SOS1*, and *SPOP* were identified in FA/miFTC. Fusion gene rearrangement was barely discovered in FA/miFTC (only one, *PAX8-PPARG*), in contrast to PTC (17.60%). The frequencies of $BRAF^{V600E}$ and *H/K/NRAS* mutations were considerably different in FA/miFTC (0% and 38.18%, respectively) and cPTC (71.43% and 1.3%, respectively). Moreover, FVPTC (25% and 47.92% for $BRAF^{V600E}$ and *H/K/NRAS* mutations, respectively) represented intermediate frequencies of those mutations relative to FA/miFTC

and cPTC. Three molecular subtypes, *BRAF*-like, *RAS*-like, and Non-*BRAF*-Non-*RAS* (NBNR), were discovered by gene expression analysis and they were more clearly stratified 180 thyroid tumors than histological classifications. The newly identified molecular subtype, NBNR, was associated with *DICER1*, *EIF1AX*, *IDH1*, *PTEN*, *SOS1*, *SPOP*, and *PAX8*-*PPARG*. Three molecular subtypes showed distinctive regulation of signaling pathways such as MAPK, p53, cell-cell communication, and metabolism pathways. The transcriptome profiles of miFTC and FA was indistinguishable from encapsulated FVPTC which is another form of benign malignancy, providing a molecular explanation for the indolent behavior of these tumors. Furthermore, the up-regulation of *ESRRA* and *PPARGC1A* which are main contributors to mitochondrial biogenesis were found in oncocytic FA/miFTC. Arm-level copy number alteration of 180 thyroid tumors was discovered by jointly-regulated block analysis and it was correlated with histological or molecular subtypes. From the aforementioned results, the current molecular understanding of WDTC is broadened and it would be lead to new diagnostic and therapeutic approaches in the future.

* This chapter was published in Plos Genetics [1].

Keywords: Thyroid carcinoma; Well-differentiated thyroid carcinoma; Papillary thyroid carcinoma; Follicular thyroid carcinoma; Massively parallel sequencing; RNA sequencing

Student number: 2014-21325

Introduction

In 2014, TCGA reported the first genomic landscape of PTC which accounts for up to 80% of WDTC [39]. In-depth molecular characterization of various subtypes of PTC was conducted and they discovered novel driver genes such as *EIF1AX*, *CHEK2* and *PPM1D*. The most remarkable result was that this type of cancer could be categorized into two major subtypes according to their molecular signatures rather than current histological features. The two molecular subtypes, $BRAF^{V600E}$-like and *RAS*-like, represented the differential signaling activation or suppression such as MAPK signaling pathway or genes related to thyroid metabolism and function.

Although their analysis conducted on a large number of specimens, it didn't fully covered WDTC, because FTC which is the second prevalent type of WDTC was leaved out from their research. It is known that distinction between follicular variant of PTC (FVPTC), FTC, and benign follicular adenoma (FA) is a huge hurdle in clinic owing to their similar cytological features [84, 106, 107]. Thus, they are often jointly referred to as follicular thyroid neoplasm (FTN) and diagnosis of these malignancies is accompanied after surgery. FTC accounts for approximately 10% of all thyroid carcinomas [76]. FTC usually have *H/K/NRAS* mutations, which could be implemented as one of the molecular markers for diagnosis [108]. However, these types of mutation are also discovered in FVPTC and FA [109, 110]. Hence, they are

not enough to fully predict the pure follicular histology and its malignant potential.

The objective of current study is to reveal molecular characteristics of follicular-patterned thyroid carcinomas comparing with PTC. As my knowledge, this is the first large scale MPS study which incorporates balanced number of FTC [especially minimally invasive FTC (miFTC)], classical PTC (cPTC), and FVPTC. Since there is no preceding large-scale RNA-seq study on miFTC and FA, it is expected that the result of this thesis will facilitate the discovery of new diagnostic and therapeutic approaches to this types of thyroid cancer.

Materials and methods

Ethics statement

The institutional review board of Seoul National University Hospital approved this study, in accordance with the Declaration of Helsinki (approved ID: H-1108-041-372). Written informed consent was obtained from each subject.

Patients and pathological diagnosis

Fresh frozen thyroid tissues after thyroid surgery were collected from March 2007 to January 2014 from 180 patients (49 men and 131 women; 47 ± 13 years of age). A total of 180 tumor tissue samples (25 FAs, 30 FTCs, 48 FVPTCs, and 77 cPTCs) and 81 adjacent normal thyroid tissues were collected. A specialized pathologist carefully reviewed the pathological slides. cPTC was defined if the tumor has well-formed papillae with fibrovascular cores and characteristic nuclear features of papillary carcinoma. FVPTC was defined as a PTC with predominantly a follicular growth pattern more than 50%, no well-formed papillae. Two subtypes of FVPTC: infiltrative FVPTC and EFVPTC was diagnosed based on the tumor border - infiltrative border or a pushing border with smooth outlines and a capsule, respectively. Capsular invasion was identified in only two cases in EFVPTC and there was not capsular invasion in the other FVPTCs. Therefore, encapsulated FVPTC was not categorized into two subgroups regarding capsular invasion. miFTC was

diagnosed if the tumor is encapsulated by capsular invasion and/or small-caliber sized angioinvasion. FA was diagnosed with no capsular invasion and angioinvasion [111].

RNA sequencing

RNA extraction from specimens was performed using the QIAcube and RNeasy Mini Kit (Qiagen, Hilden, Germany). Also, the Easy Spin RNA extraction kit (Intron, Daejeon, Korea) was used when the tissue volume was too small. RNA 6000 Nano LabChip on a 2100 Bioanalyzer (Agilent Inc., Palo Alto, CA) was used for measuring quality and concentration of extracted RNA. The RNA-seq was performed by a HiSeq 2000 platform (Illumina, San Diego, CA). The sequenced paired-end reads were aligned to hg19 human genome reference using STAR 2-pass method and Picard's MarkDuplicate (http://picard.sourceforge.net) removed polymerase chain reaction (PCR) duplicates [112]. Then, indel realignment and base quality score recalibration (BQSR) was conducted by Genome Analysis Toolkit (GATK) [113].

Driver mutation identification

MuTect identified somatic SNV from 81 samples which have adjacent normal thyroid tissues [44]. For non-matched samples, single sample mode of MuTect and GATK's HaplotypeCaller was used. Also, GATK's HaplotypeCaller was used for indel detection [113]. ANNOVAR annotated all variants called in these methods [114]. Furthermore, GATK's DepthOfCoverage was applied for

counting alternative allele of mutation hotspots in well-known oncogenes such as *BRAF* and *H*/*K*/*NRAS*.

To identify candidate driver mutations, additional filtration to variant calls were applied as follows: 1) not or rarely shown in public databases of normal individuals, such as Exome Aggregation Consortium (ExAC) [115], 1000 Genomes Project phase 3 [62], and Exome Sequencing Project 6500 (http://evs.gs.washington.edu/EVS/) [minor allele frequency (MAF) ≤ 0.0001 for ExAC and ≤ 0.01 for other databases]; 2 nonsilent SNVs (nonsynonymous and splice-site) and frameshift indels; 3) genes that were annotated in COSMIC70 or PTC dataset of TCGA project. Driver candidates in TCGA were further examined by cBioPortal for Cancer Genomics [116].

Fusion gene identification

Minimum Overlap Junction Optimizer algorithm was used for identifying fusion gene in thyroid cancer using TCGA GAF 3.0 reference (https://github.com/cband/MOJO). Further filtration steps were applied to filter false positive calls: 1) fusion genes only shown in tumor samples; 2) discordant read pairs between gene A and B ≥ 2; and 3) genomic distance between predicted coordinates ≥ 100 kb or two genes located on different chromosomes.

Transcriptome profiling and DEG analysis

Based on Ensembl gene annotation, the number of reads aligned to each gene were calculated using HTSeq [117]. DESeq2 identified DEG and following criteria were applied to rule out the false positive calls [21]; q-value < 0.05, $|Log_2FC| \geq 1$, and baseMean ≥ 100. The calculated p-values were adjusted to q-values for multiple testing using Benjamini-Hochberg method [118]. The normalized gene expression values were applied to PCA using the most variable 500 genes. The median subtracted regularized log (rlog) values were applied to the K-means clustering algorithm using cluster 3.0 for heatmap visualization [119]. To identify the significantly dysregulated pathways, Molecular Signatures Database 5.0 was used [120].

Thyroid differentiation score and ERK signature

Thyroid differentiation score (TDS) was calculated using 16 thyroid metabolism and function genes as described by TCGA study [39]. The average of median subtracted rlog values of each gene were determined as TDS.

TDS = Mean of median-subtracted rlog of 16 genes

To illustrate the MAPK signaling pathway activation, modified ERK score calculation from TCGA was conducted. The identical calculation approach which was described in TDS calculation was performed using 52 MAPK signaling pathway genes [121].

ERK score = Mean of median-subtracted rlog of 52 genes

Jointly regulated block analysis

For jointly regulated block (JRB) analysis, protein-coding genes in autosomes that have average fragment per kilobase per million (FPKM) ≥ 1.5 were selected for the analysis. After that, genes were sorted by chromosomal coordinate and applied three normalization steps as follows:

1) Log (FPKM) of gene (gene A) in i^{th} tumor sample was Z-score transformed:

$$Z_{i,A} = \frac{Log\ (FPKM) - \mu}{\sigma}$$

where μ and σ denote average and standard deviation of Log (FPKM) of 81 normal tissues.

2) Z-score of gene (gene A) in i^{th} tumor sample was Z-score transformed:

$$Z_{Z,i,A} = \frac{Z_{i,A} - \mu_i}{\sigma_i}$$

where μ_i and σ_i denote average and standard deviation of Z-score of i^{th} tumor sample.

3) Median Z-score of each chromosome arm of i^{th} tumor sample was median-centered by subtracting the median Z-score of all chromosome arms. After normalization steps, the arms with median-centered Z-score ≥ 0.5 and ≤ -0.5 were defined as arm-level duplication and deletion, respectively.

Statistical analyses of clinical data

All statistical analyses were performed using SPSS version 20.0 (IBM Co, Armonk, NY, USA). Data are presented either as frequencies (%) or as mean ± standard deviation. Comparisons of categorical variables were performed using either the Pearson's chi-square or Fisher's exact test (if the number was < 5), and the independent t-test was used for continuous variables. Adjusted *p*-values for age and sex were obtained by the binomial or multinomial logistic regression analyses for categorical variables and by either the linear regression or analysis of covariance (ANCOVA) for continuous variables. A post-hoc Bonferroni test were used to determine which groups have statistically different proportion of clinical risk factors. Statistical significance was defined as two-sided *p*-values < 0.05.

Results

Mutational profile of thyroid tumors

The average throughputs of RNA-seq of normal thyroid (3 Gb on average), FA (3.28 Gb on average), miFTC (3.09 Gb on average), FVPTC (3.04 Gb on average), and cPTC (3.13 Gb on average) are provided in Table 1-1. The mutational landscape of 180 thyroid tumors including 25 FAs, 30 miFTCs, 48 FVPTCs, and 77 cPTCs was demonstrated in Figure 1-1. Well-known thyroid cancer driver mutations such as $BRAF^{V600E}$, $HRAS^{G13R}$, and $NRAS^{Q61R}$ in addition to fusion gene driver were identified in 37.22%, 25.00%, and 12.78% of total samples, respectively. The distinctive patterns of driver mutation were found between PTC and FTN. Fusion gene rearrangements were frequently identified in PTC (17.60% and 1.82% in PTC and FTN, respectively; Fisher's exact test $p = 0.002$), while most small size mutations (SNV and indel) in driver genes except *BRAF* and *H/K/NRAS* were observed in FTN (32.73% and 0.80% in FTN and PTC, respectively; Fisher's exact test $p < 0.0001$). Only PTC harbored $BRAF^{V600E}$ and its two subtype, cPTC and FVPTC, also displayed different incidence of this mutation (71.43% and 25.00% for cPTC and FVPTC, respectively; Chi-square test $p < 0.0001$). The most recurrently altered driver genes in follicular-patterned thyroid tumors (FVPTC, miFTC, and FA) were *RAS* family genes (47.92%, 50.00%, and 24.00%, respectively) and only one cPTC had this type of alteration.

Somatic mutations in Ribonuclease III domain of *DICER1* (E1705Q, D1810H, E1813G, and E1813Q) were identified in four samples (6.67% of miFTC and 8.00% of FA; Figure 1-1). Four *DICER1* mutations were mutually exclusive with $BRAF^{V600E}$ and *H/K/NRAS* mutations. In addition, the up-regulation of *DICER1* mRNA expression with these somatic mutations were detected (Figure 1-2A). Two of four *DICER1* mutations discovered in this study were also found in TCGA study (D1810H and E1813G in TCGA-EL-A3GO and TCGA-EL-A3D5, respectively), but TCGA did not describe these mutations in their result [39]. However, they were also mutually exclusive with *BRAF* and *H/K/NRAS* mutations and accompanied increased expression of *DICER1* expression (Figure 1-2B and 1-3). Previous studies have verified the relationship between mutations in Ribonuclease III domain of *DICER1* and several types of cancer including PTC [81, 122, 123]. However, this is the first report identifying *DICER1* somatic mutation in FTN. Also, three missense mutations in *EIF1AX* (G9V, R13C, and R13L) which was defined as a driver gene in PTC were found [39, 124]. *EIF1AX* alteration was more frequently detected in FA than PTC (12.00% and 0.80%, respectively; Fisher's exact test $p = 0.015$). Mutations in *EIF1AX* were also mutually exclusive to other well-known alterations such as $BRAF^{V600E}$ and *H/K/NRAS* mutations (Figure 1-3). $IDH1^{R132C}$ and two *PTEN* missense mutations (V343E and V175A) were identified in miFTC. Also, *PTEN* loss in one FA sample was also suspected, because its mRNA expression was repressed.

In addition to aforementioned well-known cancer related genes, novel driver mutation or gene candidates were found: *SOS1*^{N233Y}, *SPOP*^{P94R}, *EZH1*^{Q571R}, *EZH1*^{Y642F}, and *STK11*^{R86fs}. Recently, *SOS1*^{N233Y} was defined as likely oncogenic mutation in several cancers and identified in uterine endometrial carcinoma, lung adenocarcinoma, and cancer cell lines (Figure 1-4) [125]. *SPOP* is recurrently altered in prostate cancer and *SPOP*^{P94R} was identified in recent study regarding FTN and benign thyroid nodules [83, 126-128]. Moreover, *EZH1*^{Q571R} and *EZH*1^{Y624F} were identified in benign thyroid nodules including autonomous thyroid adenomas and one PTC of TCGA study, respectively [39, 83, 126, 129]. In TCGA dataset, *SPOP*^{P94R} and *EZH1*^{Y642F} was mutually exclusive with *BRAF* and *H/K/NRAS* alterations (Figure 1-3). The frameshift mutation in *STK11* is frequently found in several cancer types as well as PDTC and ATC [101, 130, 131].

All fusion genes found in this study, including novel driver fusion candidate, were provided with their breakpoints (Table 1-2). They were also mutually exclusive with *BRAF*^{V600E} and *H/K/NRAS* mutations. *ETV6-NTRK3* (4.80% in PTC), *CCDC6-RET* (2.40% in PTC), *NCOA4-RET* (0.80% in PTC), *SQSTM1-NTRK1* (0.80% in PTC), *STRN-ALK* (0.80% in PTC), and *PAX8-PPARG* (0.80% and 1.82% in PTC and FTN, respectively) which have been reported in thyroid cancer were accounts for most discovered fusion genes [8]. The aberrant activations of kinase domain of *ALK*, *RET*, and *NTRK1* after fusion breakpoint which is one of the signature of oncogenic fusions were confirmed (Figure 1-5). In terms of *BRAF* fusion, *SND1-BRAF* (0.80% in

PTC) which was reported previously [39], and two novel fusions, *PICALM-BRAF* (0.80% in PTC) and *NFYA-BRAF* (0.80% in PTC) were found. Additionally, *EZR-ERBB4*, *FGFR2-KIAA1598*, *FGFR2-WARS*, *PAX8-GLIS3*, *THADA-LOC100505678*, and *RNF213-SLC26A11* were also discovered as driver fusion gene candidates. In other cancer types, *EZR-ERBB4*, *FGFR2-KIAA1598*, and *RNF213-SLC26A11* have been found previously [132-134]. The aberrant activation of *ERBB4* after fusion gene breakpoint was also detected (Figure 1-5). It is known that *THADA* rearrangement induces the activation of *IGF2BP3* expression IGF1R signaling [135]. In this study, only one sample with *THADA-LOC100505678* displayed the up-regulation of *IGF2BP3* (Figure 1-6). *PAX8-GLIS3* was also suggested as driver fusion gene owing to these two genes involve in thyroid metabolism and function [39].

Association between clinical risk factors and genetic alterations

Different combinations of clinical risk factors such as age, smoking, alcohol drinking, menopausal status, and the presence of lymphocytic thyroiditis (LT) were exhibited by each study subject. The patients were categorized into three groups to determine the potential relationship of these risk factors with diverse types of genetic alteration: 1) small size mutation, 2) fusion gene, and 3) driver-unknown (Table 1-3). The fusion gene group (39.2 ± 13.1) was younger than driver-unknown (52.3 ± 14.9) and small size mutation (47.4 ± 12.1) groups on average (p = 0.002). Moreover, young adults more frequently harbored fusion gene rearrangement (20.00% of subjects age < 45 yrs. and

6.26% of subjects age ≥ 45 yrs.; p = 0.006). The percentage of pre-menopausal women in the fusion gene group (75.00%) was higher than driver-unknown (23.53%) and small size mutation (55.00%) groups (p = 0.01). The patients with *BRAF*^{V600E} (37.31%) and fusion gene rearrangement (47.83%) showed higher frequency of LT than those with *H/K/NRAS* mutations (11.11%; p = 0.002 and p = 0.0004, respectively).

Transcriptome analysis on thyroid tumors

All study subjects were applied to K-means clustering via PCA to glance the global gene expression profile of them. PC2 axis clearly separated tumor and normal thyroid samples although one of K-means cluster owned both tumor and normal samples (Figure 1-7A). It was composed of samples with LT which was observed in 28.89% of study subjects. When the analysis repeated with only tumor samples, *BRAF*^{V600E}-positive samples without LT were distinguished from those with LT (Figure 1-7B). However, the same consequence was not identified in TCGA samples; there was no separation within *BRAF*-positive samples although 22.89% of TCGA samples were affected with LT (Figure 1-7A). I suspected the inconsistent results between two studies might be derived from the difference between gene sets used for RNA-seq analysis; UCSC and Ensembl gene sets were implemented to TCGA and this study, respectively. It was identified that 91 out of most variable 500 genes, which were used to PCA91 genes, were associated with immunoglobulin, but only four of them were found in UCSC gene set. It was

supposed to be that the immunogenic signatures derived from Ensembl gene set usage could hinder analyzing the oncogenic signaling of subsequent analysis. Therefore, only genes covered by USCS gene set were used to PCA and the decrement of immunogenic signature were confirmed. Three molecular subtypes including novel subtype, Non-*BRAF*-Non-*RAS* (NBNR), as well as *BRAF*-like and *RAS*-like, which are now well-established by TCGA were demonstrated by this approach. It was clear that the three molecular subtypes were tightly associated with type of driver gene status (Figure 1-9A). This result was repeated when the same analysis was conducted to TCGA dataset (Figure 1-9B). Because there was exceptionally high proportion of *BRAF*-positive PTC samples in TCGA, this analysis was conducted with reduced number of TGCA dataset.

Each molecular subtype was associated with different types of driver genes (Figure 1-10). *BRAF*-like was restricted to PTC and it was developed by *BRAF* (V600E, *PICALM-BRAF*, *NFYA-BRAF*, and *SND1-BRAF*) as well as several fusion genes (*NCOA4-RET*, *SQSTM1-NTRK1*, *FGFR2-WARS*, and *RNF213-SLC26A11*). There was no FTN belongs to *BRAF*-like since $BRAF^{V600E}$ and fusion gene rearrangements (except for *PAX8-PPARG*) were not found in this type of tumor. *H/K/NRAS* and fusion genes (*STRN-ALK*, *EZR-ERRB4*, and *FGFR2-KIAA1598*) were identified in *RAS*-like subtype. Moreover, *ETV6-NTRK3*, and *CCDC6-RET* were both related to *BRAF*-like and *RAS*-like subtypes. The novel molecular subtype, NBNR, was related to *DICER1*, *EIF1AX*, *IDH1*, *PTEN*, and novel driver candidates. Three fusion

gene rearrangements, *THADA-LOC100505678*, *GLIS3-PAX8*, and *PAX8-PPARG* were also associated with this novel molecular subtype.

Before investigating the relationship between molecular subtype and signaling regulation, the correlation between molecular subtype and cancer aggressiveness was measured. The significant association between lymph node metastasis (LNM) and extrathyroidal extension (ETE) which are the aggressive pathologic features, were confirmed; *BRAF*-like tumors exhibited higher incidence of LNM (37.04%) or ETE (61.73%), while other two subtypes, *RAS*-like (15.09% of LNM and 11.32% of ETE) and NBNR (0.00% of LNM and 8.70% of ETE) represented less aggressive characteristics (Figure 1-11; For both categories; Fisher's exact test $p < 0.0001$).

Two scoring approaches from TCGA study, TDS and ERK score, were conducted to measure thyroid cell differentiation and the MAPK signaling pathway regulation in three molecular subtypes. The thyroid cell differentiation and molecular subtype showed a strong negative correlation (Pearson's correlation coefficient = -0.66). Thyroid cell differentiation was most decreased in *BRAF*-like samples as TDS shown, while *RAS*-like and NBNR samples displayed preserved thyroid cell differentiation (Figure 1-11). The repression of most thyroid metabolism and function genes were identified in *BRAF*-like, in contrast to *RAS*-like and NBNR which represented similar expression levels of those genes to normal thyroid tissues. The expression levels of *DIO1*, *DIO2*, *TPO*, *SLC26A4*, and *SLC5A8* were significantly

decreased in *BRAF*-like, but *RAS*-like showed up-regulation of *DUOX1* and *DUOX2* without down-regulation of those genes. In general, there was no dysregulated genes in NBNR subtype, but oncocytic tumors which incorporate increased mitochondrial copy number due to up-regulation of *ESRRA* represented the significant down-regulation of *DIO1*, *FOXE1*, *GLIS3*, *PAX8*, and *SLC5A5* (Figure 1-12).

It is well established that MAPK signaling pathway is constitutively activated in PTC and it is closely involved in pathogenesis of this type of cancer [136]. ERK score well represented activation level of MAPK signaling pathway, Thyroid cell differentiation and the activation of MAPK signaling were strongly positively correlated (Pearson's correlation coefficient = 0.80). In consistent with TCGA study, the high activation of MAPK signaling pathway was usually found in *BRAF*-like samples, but barely in other samples which were classified as *RAS*-like or NBNR (Figure 1-11). However, there was a distinction between *RAS*-like and NBNR regarding the MAPK signaling pathway activation; NBNR did not show any up-regulation, but *RAS*-like represented increment of some genes in the MAPK signaling pathway (Figure 1-12).

Transcriptome analysis on follicular-patterned thyroid tumors

It was shown that miFTC have similar mutational profile to not only benign FA, but also encapsulated FVPTC (EFVPTC) which is also recently re-

defined as benign nodule (Figure 1-1) [137]. In contrast to EFVPTC, infiltrative FVPTC displayed was similar to cPTC with high proportion of *BRAF*^{V600E}. It is hard to precise diagnose aforementioned follicular-patterned tumors based on their histological features. Therefore, DEG analysis was performed to discover potential biomarkers which could distinguish those tumors in molecular level. At first, PCA was used to glance the difference of global gene expression patterns between EFVPTC and infiltrative FVPTC is exist. Two subtypes of FVPTC were well divided by PC1 axis (Figure 1-13). However, it was clear that molecular subtypes, rather than histological subtypes, more properly displayed the distinction between samples. In terms of miFTC and FA, they also represented highly similar transcriptome profiles and indistinguishable based on pathological classification. The PCA suggested several groups are existing in miFTC and FA in the aspect of gene expression patterns which are affected by driver gene (Figure 1-14A). Aforementioned results were further verified by DEG analysis which resulted in a negligible number of DEG among miFTC, FA, and EFVTPC (Figure 1-15).

The characteristic gene expression of oncocytic FTN

In PCA result using miFTC and FA, the one group with driver unknown samples was associated with the *ESRRA* up-regulation (Figure 1-14A and 1-14B). It was also confirmed by pathway enrichment analysis using 'chemical and genetic perturbations' database [120]; the up-regulated genes in these group was tightly involved in *ESRRA* and mitochondrial genes (Table 1-2).

This group with *ESRRA* overexpression was associated with oncocytic FTN (Fisher's exact test $p < 0.0001$); 83.33% of oncocytic FTN (Figure 1-14B).

Oncocytic FTN represents unique cytological features as increased number of mitochondria [138]. Furthermore, *PPARGC1A*, known to be key regulator of mitochondrial biogenesis, was up-regulated in those samples and its expression level was strongly positively correlated with that of *ESRRA* (Pearson correlation coefficient = 0.83 using FPKM). Also, extremely up-regulated genes in citric acid cycle (TCA cycle) were also confirmed (Figure 1-16). K-means clustering via PCA classified all the oncocytic FTNs as NBNR.

Differentially regulated signaling pathways in three molecular subtypes

Pathway enrichment analysis using dysregulated genes of each molecular subtype was performed with Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database to verify the detailed transcriptional characteristics of them [139]. Each molecular subtype was displayed with the most significantly enriched 15 pathways for up- and down-regulated DEGs, respectively. In terms of NBNR, they were separately subjected to the analysis according to type for driver gene or gene signature (*DICER1*, *EIF1AX*, *PAX8-PPARG*, and *ESRRA* up-regulation).

The remarkable up-regulated pathway in *BRAF*-like tumors were related to intercellular communications such as cell adhesion molecules (CAMs), the extracellular matrix (ECM) receptor interaction, and focal adhesion (Table 1-5). It is well reported that these signaling pathways are deeply involved in the carcinogenesis or cancer invasiveness including thyroid malignancy [140-142]. In terms of MAPK signaling pathway, it was significantly up-regulated in *BRAF*-like and *RAS*-like tumors (Table 1-5 and 1-6). However, it was only increased in *DICER1* mutated among four sub-groups of NBNR (Table 1-7-1-10). Various metabolism and calcium signaling pathways were down-regulated in *BRAF*-like, but barely in other molecular subtypes (Table 1-11-1-16).

Novel molecular subtype NBNR is composed of diverse kinds of driver gene and they had different gene expression profiles depending on the types of mutated or overexpressed gene. As mentioned earlier, up-regulated genes of *ESRRA* overexpressed tumors were significantly enriched for pathways related to TCA cycle, oxidative phosphorylation (OXPHOS), PPAR signaling, and several metabolisms (Table 1-9). Moreover, tumors with *PAX8-PPARG* rearrangement also showed increased metabolism and PPAR signaling pathways (Table 1-10). The Wnt and mTOR signaling pathways were significantly up-regulated in *DICER1* and *EIF1AX* mutated tumors, respectively (Table 1-7 and 1-8).

Arm-level CNA profile of thyroid tumors

To illustrate the arm-level CNA profile of 180 thyroid tumors, JRB analysis was performed. In cancer genome, JRB and CNA are significantly correlated [32]. The co-localized over- and under-expressed chromosome arms denote arm-level copy number gain and loss, respectively. The aberration of chromosome arms was successfully predicted which represent arm-level amplification and deletion (Figure 1-17). The arm-level CNA profile of 180 thyroid tumors was displayed in Figure 1-18. The high percentage of arm-level CNA of follicular-patterned thyroid tumors, miFTC, EFVPTC, infiltrative FVPTC, were found (Figure 1-19A). Meanwhile, cPTC had the lowest proportion of arm-level CNA relative to even benign FA. The recurrent gain of chromosome 12 in FTN was detected in parallel with recent report which described frequent amplification of this chromosome in indolent thyroid tumors [39]. From the point of view of molecular subtype, *RAS*-like harbored more frequent arm-level deletion than *BRAF*-like and NBNR (Figure 1-19B). The most recurrent type of chromosome alteration was 22q deletion in *RAS*-like as recent reports (Chi-square test $p < 0.001$; Figure 1-19C). Furthermore, the amplification of 18p was more frequently found than *RAS*-like or NBNR although it was not statistically significant (Chi-square test $p = 0.051$).

Table 1-1. The average RNA sequencing summary of each tissue type.

Tissue type	Total reads	Uniquely mapped reads	Uniquely mapped reads (%)
Normal thyroid	29,970,798	25,829,904	85.54
FA	32,741,745	28,553,641	87.03
miFTC	30,885,716	26,767,556	86.46
FVPTC	30,445,888	26,115,318	85.71
cPTC	31,318,258	26,779,479	85.48

Table 1-2. The list of fusion gene drivers identified from 180 study subjects [1].

Fusion gene	Chr (5')	Strand (5')	Breakpoint (5')	Breakpoint region (5')	Chr (3')	Strand (3')	Breakpoint (3')	Breakpoint region (3')	Distance
PAX8-PPARG	chr2	-	113,994,177	UTR-3	chr3	+	12,421,202	CDS	CTX
PAX8-PPARG	chr2	-	113,992,970	UTR-3	chr3	+	12,421,202	CDS	CTX
EZR-ERBB4	chr6	-	159,190,357	CDS	chr2	-	212,488,769	CDS	CTX
FGFR2-KIAA1598	chr10	-	123,243,211	CDS	chr10	-	118,713,721	CDS	4,529,490
THADA-LOC100505678	chr2	-	43,506,903	CDS	chr5	+	111,065,187	ncRNA	CTX
ETV6-NTRK3	chr12	+	12,006,495	CDS	chr15	-	88,576,276	CDS	CTX
SQSTM1-NTRK1	chr5	+	179,252,226	CDS	chr1	+	156,845,311	CDS	CTX
PICALM-BRAF	chr11	-	85,685,750	CDS	chr7	-	140,487,384	CDS	CTX
FGFR2-WARS	chr10	-	123,243,211	CDS	chr14	-	100,828,258	CDS	CTX
NCOA4-RET	chr10	+	51,582,272	CDS	chr10	+	43,612,031	CDS	7,970,241
SND1-BRAF	chr7	+	127,361,454	CDS	chr7	-	140,487,384	CDS	1,312,5930
BRAF-NFYA	chr7	-	140,482,820	CDS	chr6	+	41,062,134	CDS	CTX
CCDC6-RET	chr10	-	61,665,879	UTR-3	chr10	+	43,612,031	CDS	18,053,848
PAX8-GLIS3	chr2	-	114,035,946	UTR-3	chr9	-	4,118,881	CDS	CTX

Table 1-2. Continued.

Fusion gene	Chr (5')	Strand (5')	Breakpoint (5')	Breakpoint region (5')	Chr (3')	Strand (3')	Breakpoint (3')	Breakpoint region (3')	Distance
RNF213-SLC26A11	chr17	+	78,324,196	CDS	chr17	+	78,221,928	CDS	102,268
STRN-ALK	chr2	-	37,143,220	UTR-3	chr2	-	29,446,394	CDS	7,696,826

Abbreviations: Chr, chromosome; CDS, coding DNA sequence; CTX, inter-chromosomal translocation; UTR, untranslated region.

[a]Distance denotes the chromosomal distance between two genes.

Table 1-3. Comparison of clinical risk factors among the groups with different types of mutation [1].

Variable	Driver-unknown	Small size mutation				Fusion gene	p [a]	p [b]
		Total	*BRAF*	*H/K/NRAS*	Others			
N	26	131	67	45	19	23		
Age	52.3 ± 14.9[c]	47.3 ± 12.1[c]	47.2 ± 12.0	48.1 ± 12.5	45.8 ± 12.1	39.2 ± 13.1	**0.002**	**0.010**
Sex (female)	21 (80.77)	93 (70.99)	50 (74.62)	28 (62.22)	15 (78.95)	17 (73.91)	0.587	0.425
Pre-menopause (regular)	4/17 (23.53)[c]	44/80 (55.00)	20/38 (52.63)	14/27 (51.85)	10/15 (66.67)	12/16 (75.00)	**0.010**	**0.038**
Smoking (current + ex)	3 (11.54)	20/130 (15.38)	9 (13.43)	7/44 (15.91)	4 (21.05)	1 (4.35)	0.431	0.578
Male	1/5 (20.00)	16/38 (42.11)	8/17 (47.05)	6/17 (35.29)	2/4 (50.00)	1/6 (16.66)	0.446	0.659
Female	2/21 (9.52)	4/92 (4.35)	1/50 (2.00)	1/27 (3.70)	2/15 (13.33)	0/17 (0.00)	0.400	0.172
Drinking (current)	7 (26.92)	40/130 (30.77)	16 (23.88)	16/44 (36.36)	8 (42.11)	7 (30.43)	0.926	0.482
Lymphocytic thyroiditis	7 (26.92)	34 (25.95)	25 (37.31)	5 (11.11)[c,d]	4 (21.05)	11 (47.83)	0.100	**0.008**

[a] p-value for comparison among Driver-unknown, Total, and Fusion gene mutation groups.

[b] p-value for comparison among Driver-unknown, *BRAF*, *H/K/NRAS*, Others, and Fusion gene mutation groups.

[c] Significantly different from "Fusion gene mutation" group. ($p < 0.05$ for post-hoc Bonferroni test)

[d] Significantly different from "*BRAF*" group. ($p = 0.002$ for post-hoc Bonferroni test)

Table 1-4. The result of gene set enrichment analysis using up-regulated genes in *ESRRA* overexpressed thyroid tumors [1].

Gene Set Name (Chemical and genetic perturbations)	Description	*p*-value	*q*-value
ESRRA TARGETS UP	Genes up-regulated by *ESRRA* [GeneID=2101] only.	2.1E-205	7.14E-202
ESRRA TARGETS	Genes regulated by *ESRRA* [GeneID=2101] in MCF-7 cells (breast cancer).	2.04E-194	3.45E-191
HUMAN MITODB 6 2002	Mitochondrial genes; based on literature and sequence annotation resources and converted to Affymetrix HG-U133A probe sets	6.09E-188	6.89E-185
MITOCHONDRIA	Mitochondrial genes	4.95E-170	4.2E-167
MITOCHONDRIA GENE MODULE	Genes that comprise the mitochondria gene module	5.96E-141	4.05E-138

Table 1-5. The top 15 significantly up-regulated pathways in *BRAF*-like thyroid tumors.

Pathway	*p*-value	*q*-value
ECM RECEPTOR INTERACTION	7.66E-25	1.42E-22
CELL ADHESION MOLECULES	2.5E-20	2.32E-18
FOCAL ADHESION	1.52E-19	9.40E-18
CYTOKINE-CYTOKINE RECEPTOR INTERACTION	1.03E-17	4.81E-16
PATHWAYS IN CANCER	7.42E-16	2.76E-14
LEISHMANIA INFECTION	1.55E-14	4.82E-13
COMPLEMENT AND COAGULATION CASCADES	2.01E-12	5.35E-11
ARRHYTHMOGENIC RIGHT VENTRICULAR CARDIOMYOPATHY	8.95E-12	2.08E-10
P53 SIGNALING PATHWAY	2.98E-11	6.15E-10
VIRAL MYOCARDITIS	8.41E-10	1.56E-08
DILATED CARDIOMYOPATHY	1.64E-9	2.77E-08
AXON GUIDANCE	2.49E-9	3.86E-08
REGULATION OF ACTIN CYTOSKELETON	4.79E-9	6.85E-08
MAPK SIGNALING PATHWAY	3.53E-8	4.70E-07
HYPERTROPHIC CARDIOMYOPATHY	5.68E-8	7.04E-07

Table 1-6. The top 15 significantly up-regulated pathways in *RAS*-like thyroid tumors.

Pathway	*p*-value	*q*-value
P53 SIGNALING PATHWAY	1.14E-10	2.12E-08
ECM RECEPTOR INTERACTION	1.51E-8	1.40E-06
PATHWAYS IN CANCER	5.95E-8	3.69E-06
MELANOMA	8.12E-6	3.15E-04
CHRONIC MYELOID LEUKEMIA	9.78E-6	3.15E-04
CELL ADHESION MOLECULES	1.02E-5	3.15E-04
MAPK SIGNALING PATHWAY	2.19E-5	5.74E-04
SMALL CELL LUNG CANCER	2.47E-5	5.74E-04
PURINE METABOLISM	3.97E-5	7.60E-04
DILATED CARDIOMYOPATHY	4.46E-5	7.60E-04
FOCAL ADHESION	4.49E-5	7.60E-04
CYTOKINE-CYTOKINE RECEPTOR INTERACTION	1.04E-4	1.61E-03
TYPE II DIABETES MELLITUS	1.14E-4	1.64E-03
ARRHYTHMOGENIC RIGHT VENTRICULAR CARDIOMYOPATHY	1.28E-4	1.70E-03
CARDIAC MUSCLE CONTRACTION	1.7E-4	2.11E-03

Table 1-7. The top 15 significantly up-regulated pathways in *DICER1* mutated thyroid tumors.

Pathway	*p*-value	*q*-value
PATHWAYS IN CANCER	1.68E-14	3.13E-12
MAPK SIGNALING PATHWAY	1.53E-9	1.43E-07
SMALL CELL LUNG CANCER	8.41E-9	5.21E-07
FOCAL ADHESION	4.31E-8	2.00E-06
AXON GUIDANCE	7.27E-7	2.71E-05
PANCREATIC CANCER	2.66E-6	8.26E-05
WNT SIGNALING PATHWAY	3.43E-6	9.12E-05
ECM RECEPTOR INTERACTION	1.06E-5	2.46E-04
ABC TRANSPORTERS	1.73E-5	3.57E-04
CHRONIC MYELOID LEUKEMIA	3.61E-5	6.72E-04
INSULIN SIGNALING PATHWAY	5.97E-5	1.01E-03
ACUTE MYELOID LEUKEMIA	1.04E-4	1.61E-03
PROSTATE CANCER	1.29E-4	1.81E-03
SPHINGOLIPID METABOLISM	1.36E-4	1.81E-03
LYSOSOME	1.46E-4	1.81E-03

Table 1-8. The top 15 significantly up-regulated pathways in *EIF1AX* mutated thyroid tumors.

Pathway	*p*-value	*q*-value
FOCAL ADHESION	4.25E-13	7.90E-11
CALCIUM SIGNALING PATHWAY	9.16E-10	8.52E-08
PHOSPHATIDYLINOSITOL SIGNALING SYSTEM	1.06E-8	6.56E-07
MELANOGENESIS	8.28E-8	3.85E-06
PATHWAYS IN CANCER	2.97E-7	1.10E-05
ECM RECEPTOR INTERACTION	5.66E-7	1.76E-05
ARRHYTHMOGENIC RIGHT VENTRICULAR CARDIOMYOPATHY	7.72E-6	1.80E-04
VEGF SIGNALING PATHWAY	7.72E-6	1.80E-04
SMALL CELL LUNG CANCER	1.26E-5	2.61E-04
GAP JUNCTION	1.77E-5	3.28E-04
GNRH SIGNALING PATHWAY	3.08E-5	5.22E-04
MTOR SIGNALING PATHWAY	3.53E-5	5.47E-04
VASCULAR SMOOTH MUSCLE CONTRACTION	5.74E-5	7.71E-04
CHEMOKINE SIGNALING PATHWAY	6.18E-5	7.71E-04
ACUTE MYELOID LEUKEMIA	6.22E-5	7.71E-04

Table 1-9. The top 15 significantly up-regulated pathways in *ESRRA* overexpressed thyroid tumors.

Pathway	*p*-value	*q*-value
OXIDATIVE PHOSPHORYLATION	3.55E-81	6.61E-79
PARKINSON'S DISEASE	8.8E-77	8.19E-75
ALZHEIMER'S DISEASE	1.23E-71	7.64E-70
HUNTINGTON'S DISEASE	6.1E-71	2.83E-69
CITRATE CYCLE	2.96E-34	1.10E-32
CARDIAC MUSCLE CONTRACTION	1.75E-19	5.43E-18
PROPANOATE METABOLISM	5.47E-19	1.45E-17
GLYCOLYSIS GLUCONEOGENESIS	1.85E-18	4.29E-17
PYRUVATE METABOLISM	3.52E-17	7.27E-16
VALINE LEUCINE AND ISOLEUCINE DEGRADATION	5.86E-15	1.09E-13
AMINOACYL TRNA BIOSYNTHESIS	3.73E-14	6.31E-13
GLYOXYLATE AND DICARBOXYLATE METABOLISM	7.55E-13	1.17E-11
BUTANOATE METABOLISM	9.78E-13	1.40E-11
FATTY ACID METABOLISM	1.18E-12	1.57E-11
PEROXISOME	1.07E-9	1.32E-08

Table 1-10. The top 15 significantly up-regulated pathways in *PAX8-PPARG* positive thyroid tumors.

Pathway	*p*-value	*q*-value
RIBOSOME	3.35E-74	6.23E-72
HUNTINGTON'S DISEASE	2.43E-43	2.26E-41
OXIDATIVE PHOSPHORYLATION	3.61E-42	2.24E-40
PARKINSON'S DISEASE	5.6E-40	2.36E-38
ALZHEIMER'S DISEASE	6.34E-40	2.36E-38
PROTEASOME	2.94E-15	9.11E-14
PURINE METABOLISM	3.79E-15	1.01E-13
CARDIAC MUSCLE CONTRACTION	4.48E-15	1.04E-13
PYRIMIDINE METABOLISM	4.26E-13	8.81E-12
P53 SIGNALING PATHWAY	3.68E-12	6.84E-11
PATHWAYS IN CANCER	8.65E-8	1.46E-06
ARGININE AND PROLINE METABOLISM	1.02E-7	1.57E-06
PEROXISOME	1.55E-7	2.22E-06
SPLICEOSOME	6.97E-7	9.25E-06
PPAR SIGNALING PATHWAY	1.68E-6	2.09E-05

Table 1-11. The top 15 significantly down-regulated pathways in *BRAF*-like thyroid tumors.

Pathway	*p*-value	*q*-value
PATHWAYS IN CANCER	7.05E-08	1.31E-05
HEDGEHOG SIGNALING PATHWAY	2.63E-07	2.45E-05
WNT SIGNALING PATHWAY	1.66E-06	1.03E-04
MAPK SIGNALING PATHWAY	3.38E-06	1.57E-04
GLYCINE SERINE AND THREONINE METABOLISM	2.68E-05	9.97E-04
VASCULAR SMOOTH MUSCLE CONTRACTION	6.05E-05	1.72E-03
TGF BETA SIGNALING PATHWAY	6.46E-05	1.72E-03
TYROSINE METABOLISM	1.21E-04	2.81E-03
METABOLISM OF XENOBIOTICS BY CYTOCHROME P450	1.62E-04	3.35E-03
DRUG METABOLISM CYTOCHROME P450	1.90E-04	3.53E-03
B-CELL RECEPTOR SIGNALING PATHWAY	2.37E-04	3.93E-03
GLYCEROLIPID METABOLISM	2.53E-04	3.93E-03
ARGININE AND PROLINE METABOLISM	4.01E-04	5.42E-03
BASAL CELL CARCINOMA	4.37E-04	5.42E-03
STEROID HORMONE BIOSYNTHESIS	4.37E-04	5.42E-03

Table 1-12. The top 15 significantly down-regulated pathways in *RAS*-like thyroid tumors.

Pathway	*p*-value	*q*-value
CYTOKINE-CYTOKINE RECEPTOR INTERACTION	8.75E-20	1.63E-17
CHEMOKINE SIGNALING PATHWAY	1.33E-18	1.24E-16
PRIMARY IMMUNODEFICIENCY	1.02E-17	6.34E-16
B-CELL RECEPTOR SIGNALING PATHWAY	4.08E-15	1.90E-13
T-CELL RECEPTOR SIGNALING PATHWAY	1.16E-13	4.31E-12
FOCAL ADHESION	6.23E-13	1.93E-11
HEMATOPOIETIC CELL LINEAGE	9.68E-13	2.57E-11
CELL ADHESION MOLECULES	4.29E-12	9.98E-11
PATHWAYS IN CANCER	4.90E-12	1.01E-10
LEUKOCYTE TRANSENDOTHELIAL MIGRATATION	7.61E-11	1.41E-09
REGULATION OF ACTIN CYTOSKELETON	1.58E-10	2.67E-09
COMPLEMENT AND COAGULATION CASCADES	2.22E-09	3.44E-08
MAPK SIGNALING PATHWAY	5.50E-09	7.87E-08
NATURAL KILLER CELL MEDIATED CYTOTOXICITY	5.92E-09	7.87E-08
JAK-STAT SIGNALING PATHWAY	2.90E-08	3.60E-07

Table 1-13. The top 15 significantly down-regulated pathways in *DICER1* mutated thyroid tumors.

Pathway	*p*-value	*q*-value
CYTOKINE-CYTOKINE RECEPTOR INTERACTION	1.10E-16	2.04E-14
CELL ADHESION MOLECULES	6.03E-15	5.61E-13
COMPLEMENT AND COAGULATION CASCADES	3.98E-11	2.37E-09
FOCAL ADHESION	5.10E-11	2.37E-09
B-CELL RECEPTOR SIGNALING PATHWAY	1.41E-10	4.69E-09
HEMATOPOIETIC CELL LINEAGE	1.55E-10	4.69E-09
PATHWAYS IN CANCER	1.76E-10	4.69E-09
CHEMOKINE SIGNALING PATHWAY	5.83E-10	1.35E-08
MAPK SIGNALING PATHWAY	7.51E-10	1.55E-08
PRIMARY IMMUNODEFICIENCY	4.10E-09	7.62E-08
LEUKOCYTE TRANSENDOTHELIAL MIGRATATION	8.96E-08	1.52E-06
JAK-STAT SIGNALING PATHWAY	1.14E-07	1.77E-06
VEGF SIGNALING PATHWAY	1.56E-07	2.23E-06
T-CELL RECEPTOR SIGNALING PATHWAY	1.21E-06	1.60E-05
AXON GUIDANCE	1.68E-06	2.08E-05

Table 1-14. The top 15 significantly down-regulated pathways in *EIF1AX* mutated thyroid tumors.

Pathway	*p*-value	*q*-value
CELL ADHESION MOLECULES	1.52E-13	2.82E-11
HEMATOPOIETIC CELL LINEAGE	4.54E-13	4.22E-11
MAPK SIGNALING PATHWAY	5.31E-11	3.29E-09
COMPLEMENT AND COAGULATION CASCADES	7.82E-11	3.64E-09
ANTIGEN PROCESSING AND PRESENTATION	1.25E-10	4.08E-09
LEISHMANIA INFECTION	1.32E-10	4.08E-09
VIRAL MYOCARDITIS	1.56E-10	4.14E-09
CYTOKINE-CYTOKINE RECEPTOR INTERACTION	3.62E-10	8.41E-09
INTESTINAL IMMUNE NETWORK FOR IGA PRODUCTION	9.30E-09	1.92E-07
ALLOGRAFT REJECTION	2.39E-08	4.45E-07
GRAFT VERSUS HOST DISEASE	5.51E-08	9.32E-07
PATHWAYS IN CANCER	6.43E-08	9.96E-07
TYPE I DIABETES MELLITUS	8.10E-08	1.14E-06
ASTHMA	8.54E-08	1.14E-06
SYSTEMIC LUPUS ERYTHEMATOSUS	2.84E-07	3.53E-06

Table 1-15. The top 15 significantly down-regulated pathways in *ESRRA* overexpressed thyroid tumors.

Pathway	*p*-value	*q*-value
PRIMARY IMMUNODEFICIENCY	6.73E-22	1.25E-19
CYTOKINE-CYTOKINE RECEPTOR INTERACTION	4.39E-14	4.08E-12
CHEMOKINE SIGNALING PATHWAY	3.45E-13	2.14E-11
T-CELL RECEPTOR SIGNALING PATHWAY	5.86E-13	2.73E-11
COMPLEMENT AND COAGULATION CASCADES	1.52E-12	5.66E-11
HEMATOPOIETIC CELL LINEAGE	2.38E-11	7.39E-10
B-CELL RECEPTOR SIGNALING PATHWAY	1.01E-10	2.69E-09
FOCAL ADHESION	1.53E-09	3.56E-08
CELL ADHESION MOLECULES	3.13E-08	6.46E-07
MAPK SIGNALING PATHWAY	2.54E-06	4.73E-05
NATURAL KILLER CELL MEDIATED CYTOXICITY	4.83E-06	8.17E-05
JAK-STAT SIGNALING PATHWAY	1.20E-05	1.85E-04
AXON GUIDANCE	3.06E-05	4.38E-04
FC GAMMA R MEDIATED PHAGOCYTOSIS	5.49E-05	7.30E-04
TOLL LIKE RECEPTOR SIGNALING PATHWAY	7.27E-05	9.02E-04

Table 1-16. The top 15 significantly down-regulated pathways in *PAX8-PPARG* positive thyroid tumors.

Pathway	*p*-value	*q*-value
PATHWAYS IN CANCER	1.60E-10	2.61E-08
AXON GUIDANCE	3.21E-10	2.61E-08
COMPLEMENT AND COAGULATION CASCADES	4.21E-10	2.61E-08
WNT SIGNALING PATHWAY	3.77E-09	1.76E-07
CYTOKINE-CYTOKINE RECEPTOR INTERACTION	7.69E-09	2.86E-07
CELL ADHESION MOLECULES	2.27E-07	7.04E-06
MAPK SIGNALING PATHWAY	7.66E-07	2.04E-05
CALCIUM SIGNALING PATHWAY	1.37E-06	3.19E-05
HYPERTROPHIC CARDIOMYOPATHY	4.11E-06	8.49E-05
DILATED CARDIOMYOPATHY	8.43E-06	1.57E-04
ARRHYTHMOGENIC RIGHT VENTRICULAR	1.18E-05	1.99E-04
ABC TRANSPORTERS	1.56E-05	2.42E-04
MELANOGENESIS	2.11E-05	3.02E-04
ECM RECEPTOR INTERACTION	2.67E-05	3.34E-04
FOCAL ADHESION	2.70E-05	3.34E-04

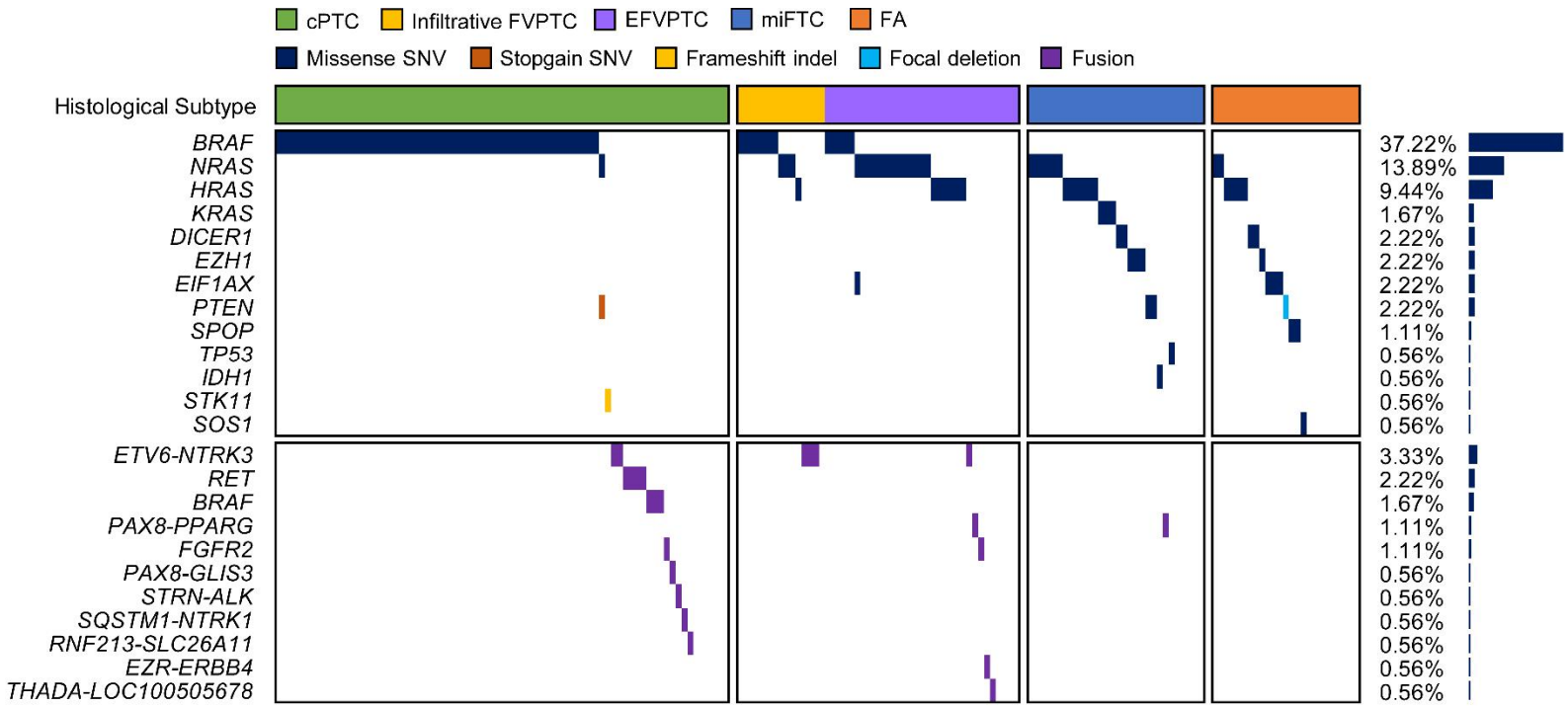

Figure 1-1. The mutational landscape of 180 thyroid tumors [1]. Each column represents an individual sample.

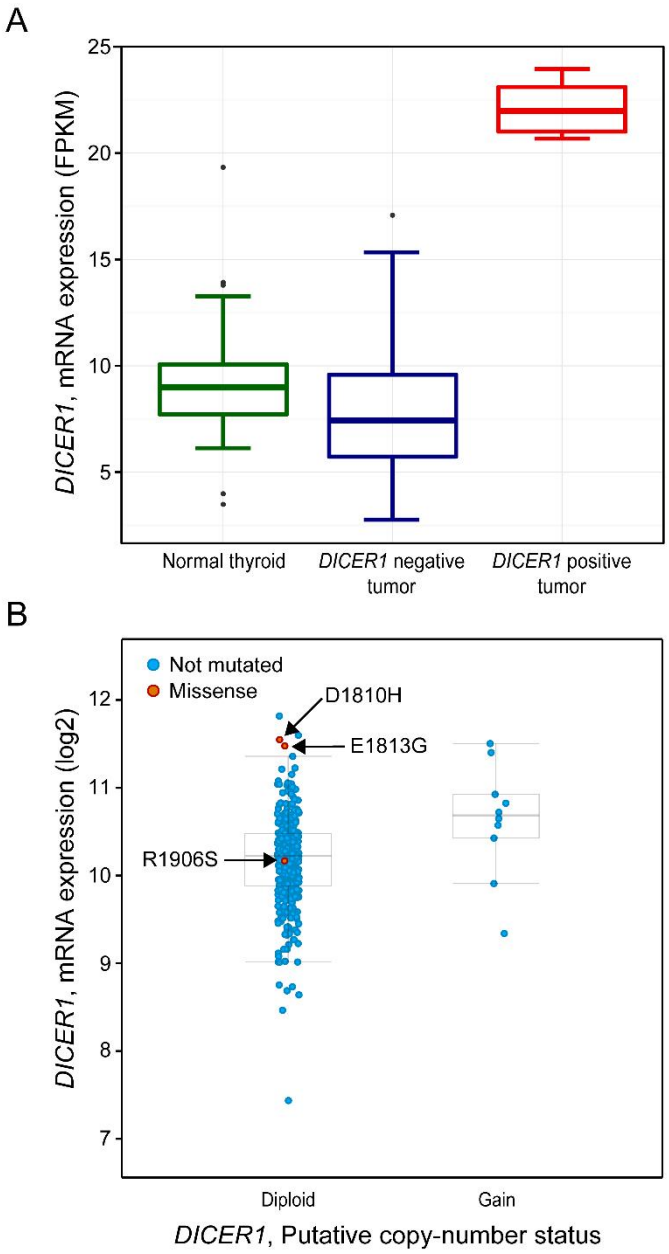

Figure 1-2. The up-regulation of *DICER1* mRNA expression with driver mutations [1]. The *DICER1* mRNA expression profile of samples in A) the current study. B) TCGA study. The plot with TCGA samples were generated by cBioPortal for cancer genomics [116].

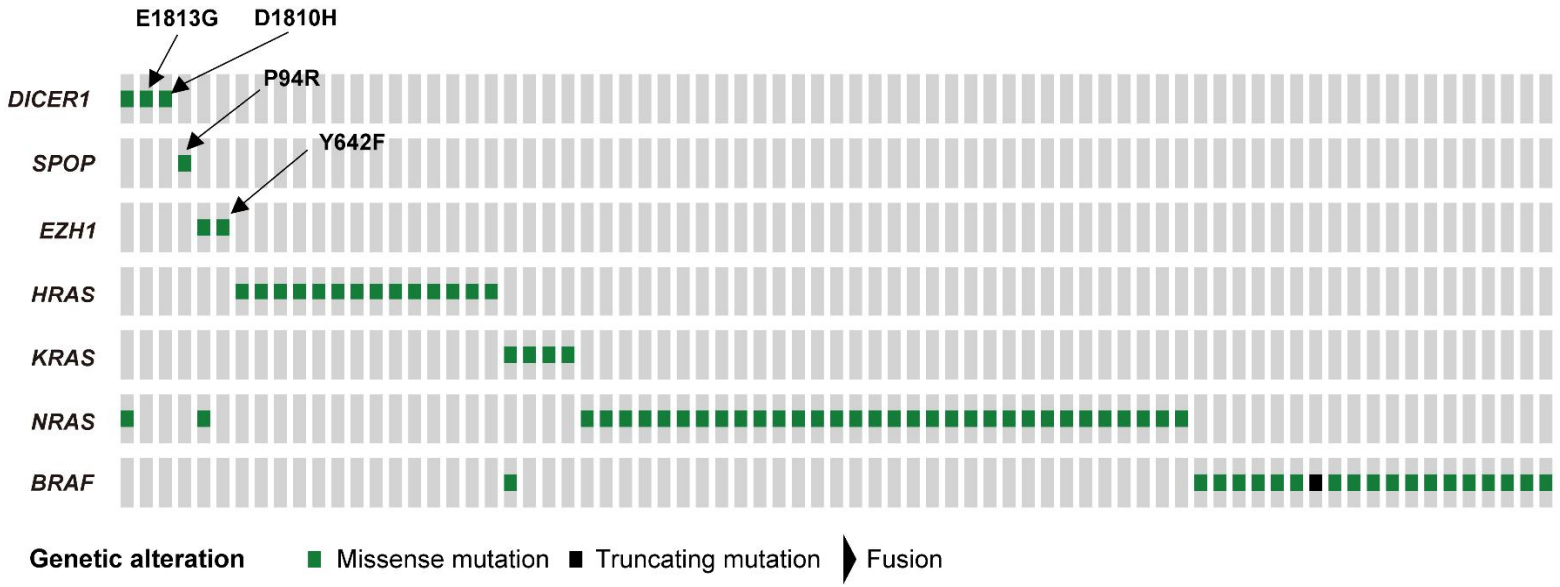

Figure 1-3. The driver candidates of thyroid tumor [1]. The distribution of driver mutation candidates in TCGA dataset. The samples on the right matrix were not displayed since they are either *BRAF* positive or their driver mutations were not identified.

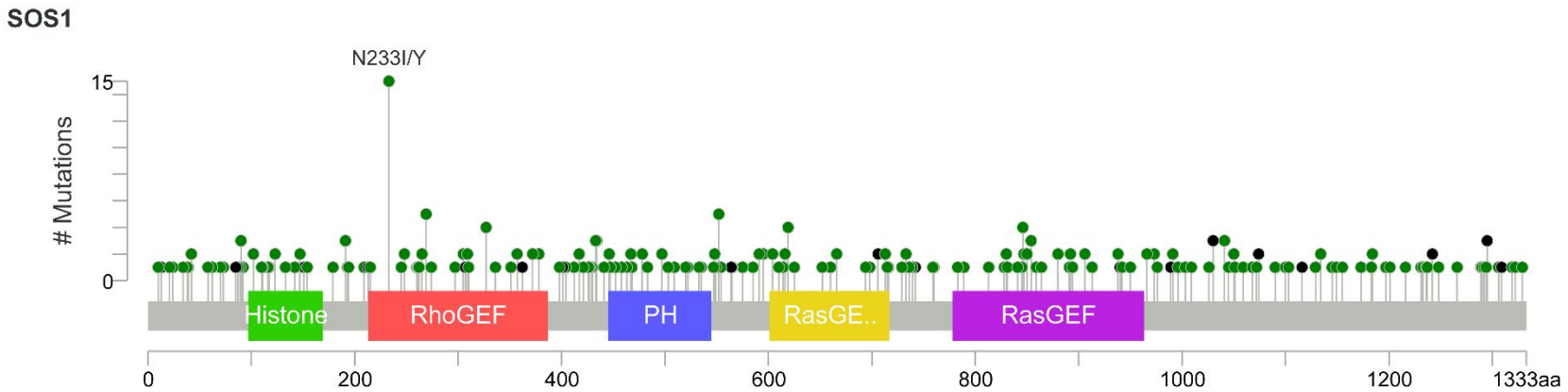

Figure 1-4. The lollipop plot displays the incidence of *SOS1*^{N233Y} mutation in TCGA dataset. The number of somatic mutations in *SOS1* across all cancer type of TCGA dataset was shown. The black and green circles indicate truncating and missense mutations, respectively.

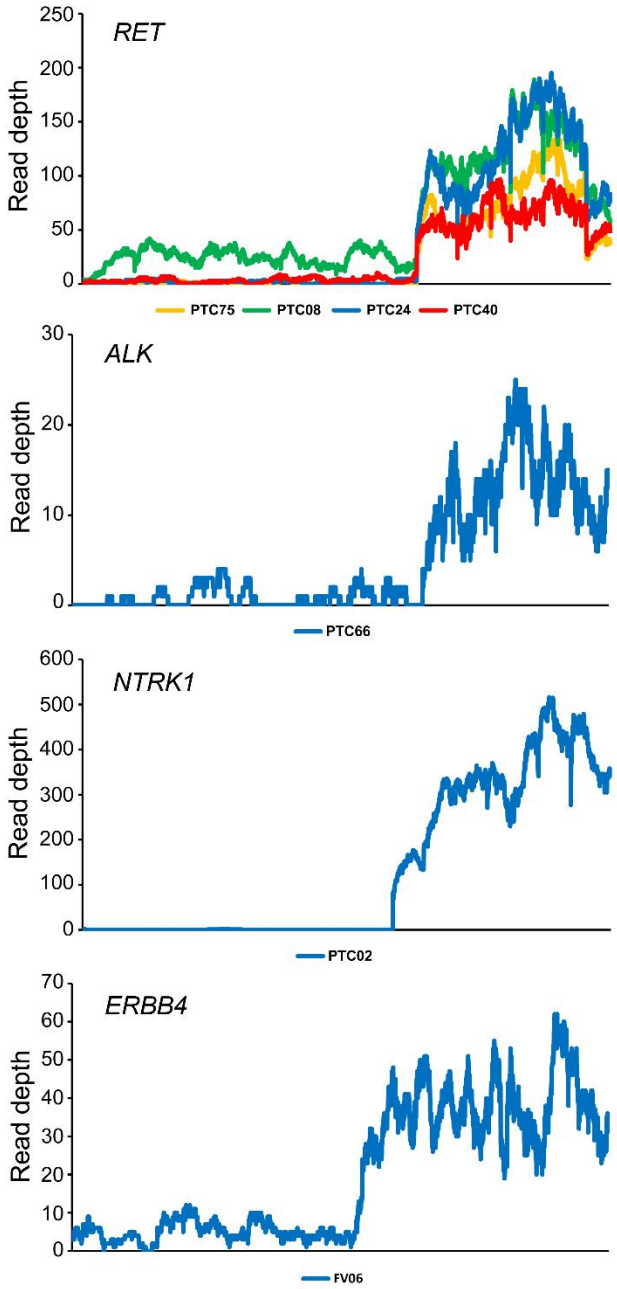

Figure 1-5. The aberrant over-expression of kinase fusion genes [1]. Each plot displays the base resolution expression level of gene.

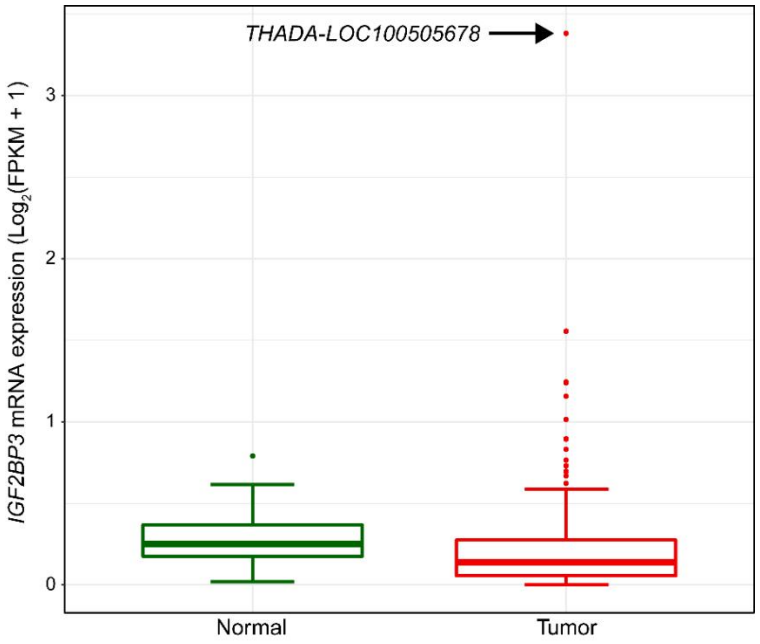

Figure 1-6. The up-regulation of *IGF2BP3* in *THADA* fusion positive tumor. The arrow indicates a sample with *THADA*-LOC100505678 fusion gene.

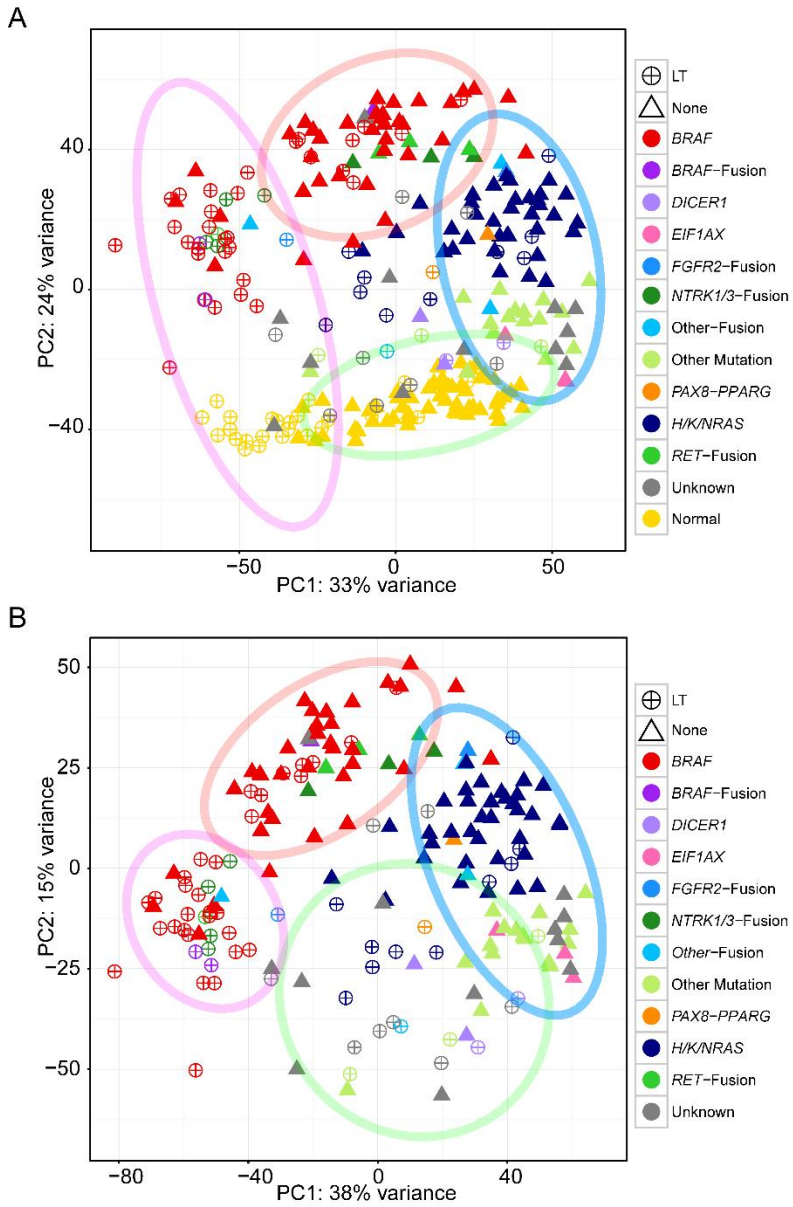

Figure 1-7. The result of PCA with Ensemble gene set [1]. A) The analysis using normal thyroid and tumor samples. B) The analysis using tumor samples. The most variable 500 genes were used. The LT status and the driver mutation of each sample were represented by the shape and the color, respectively. Each cluster was represented by a 95.00% confidence ellipse.

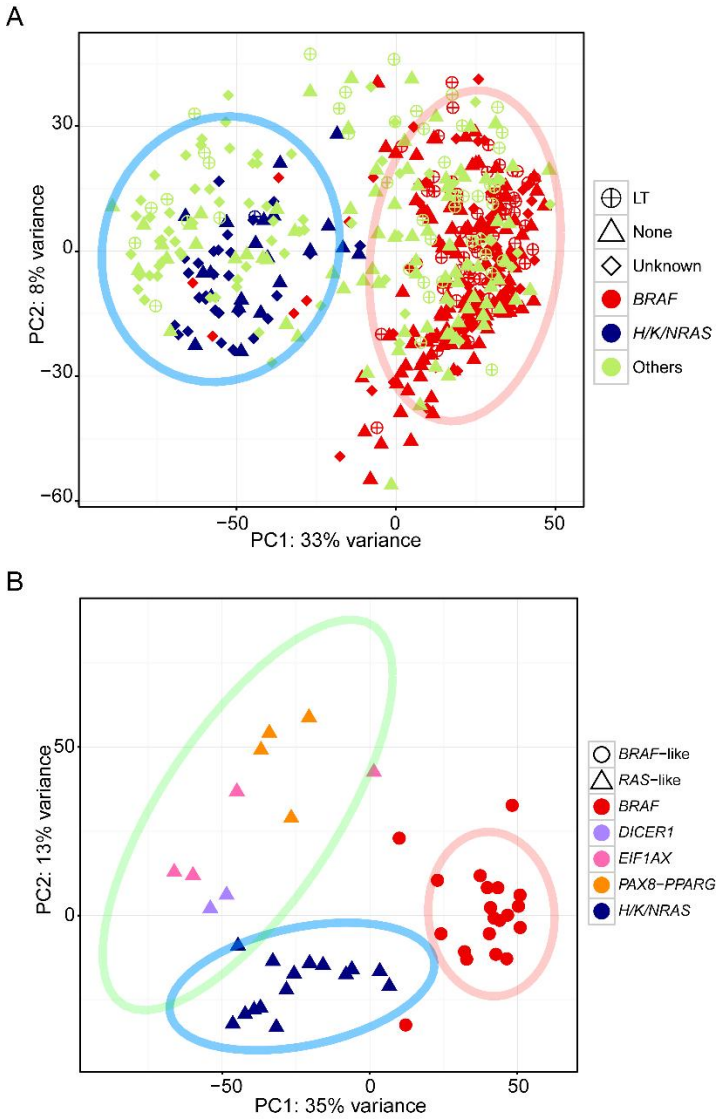

Figure 1-8. The result of PCA with TCGA dataset [1]. A) The analysis using all TCGA tumor samples. The LT status and the driver mutation of each sample were represented by the shape and the color, respectively. B) The analysis using partial TCGA tumor samples. The molecular subtype and the driver mutation of each sample were represented by the shape and the color, respectively. The most variable 500 genes were used. Each cluster was represented by a 95.00% confidence ellipse.

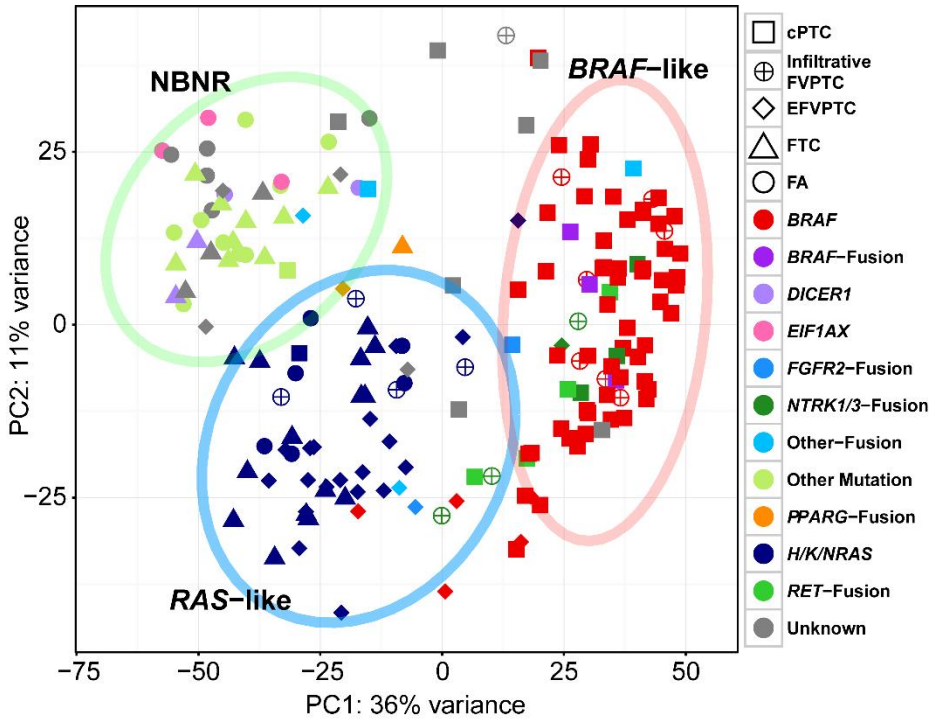

Figure 1-9. The molecular subtype classification of 180 thyroid tumors [1]. The histological subtype and the driver mutation of each sample were represented by the shape and the color, respectively. The most variable 500 genes were used. Each cluster was represented by a 95.00% confidence ellipse.

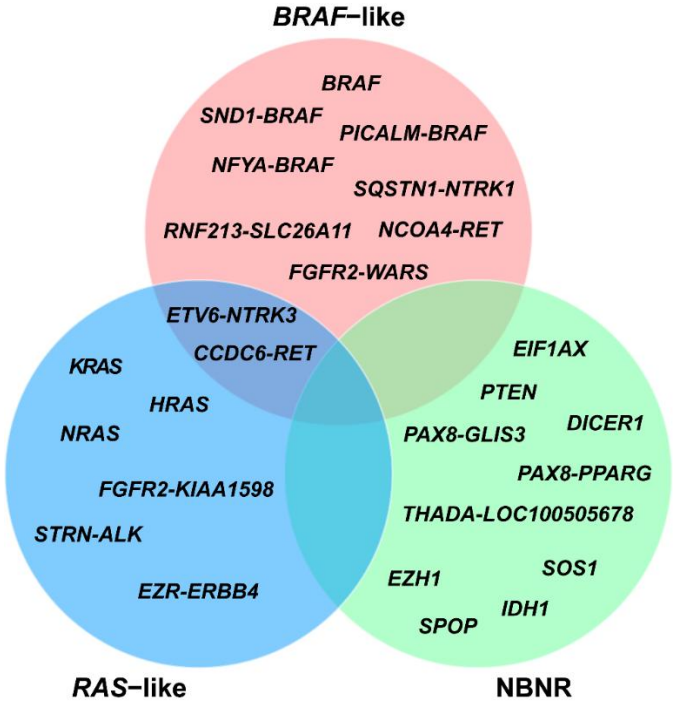

Figure 1-10. The association between driver genes and molecular subtypes [1].

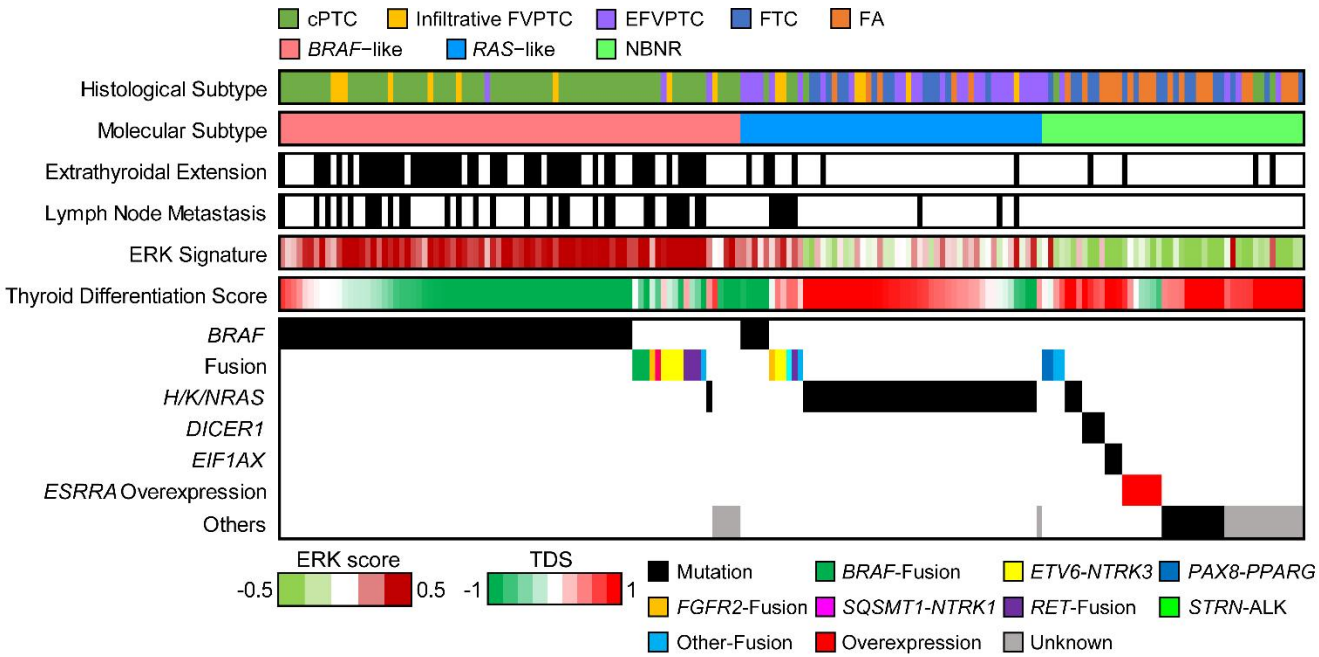

Figure 1-11. The association between three molecular subtypes and the clinical characteristics or signaling pathways [1]. The Histological subtype, the molecular subtype, the presence of ETE, the presence of LNM, ERK score, TDS, and driver gene in individual sample were displayed. 180 tumors were sorted by molecular subtype, driver gene, and high to low TDS.

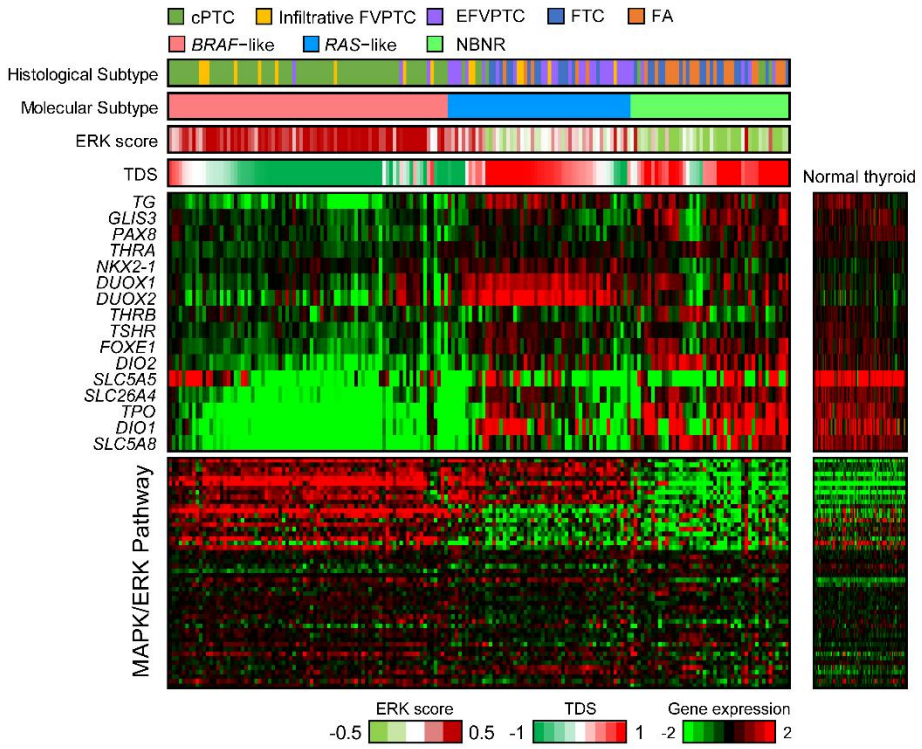

Figure 1-12. The expression pattern of 16 thyroid metabolism and function genes and the MAPK signaling pathway [1]. Genes in tumor (middle) and normal (right) samples were displayed as heatmap. Samples were sorted in the same manner as Figure 1-11. Genes were clustered by K-means clustering algorithm (K = 3).

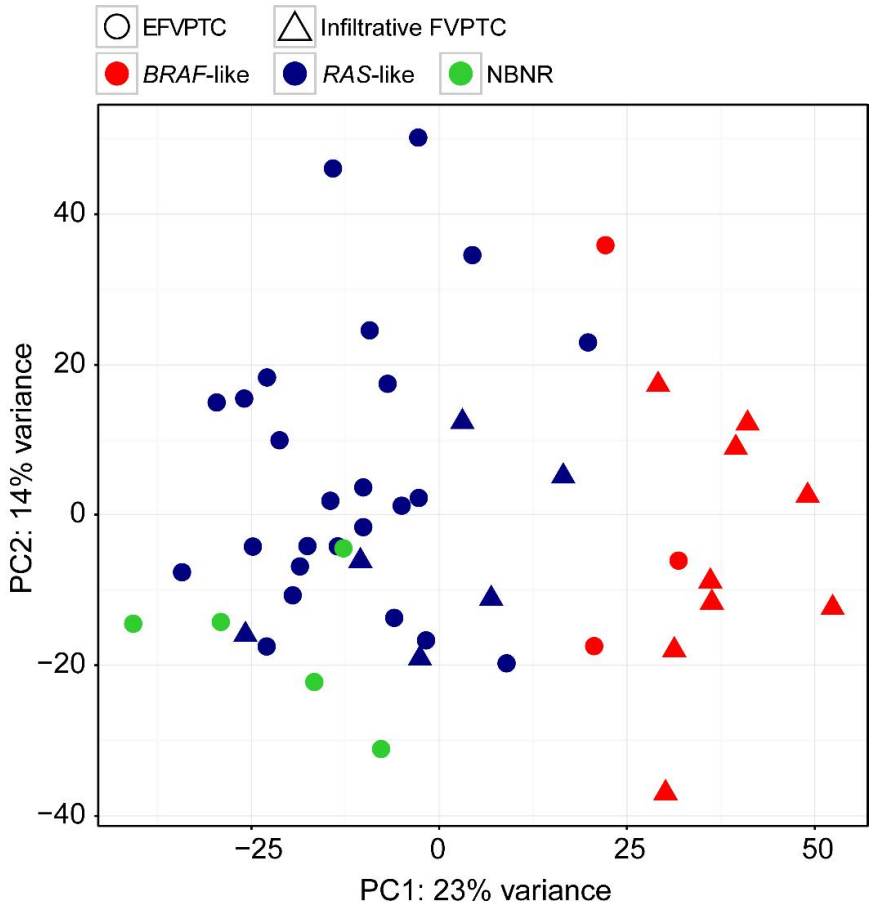

Figure 1-13. The gene expression analysis on FVPTC [1]. The result of PCA on FVPTC. The histological subtype and molecular subtype of each sample were represented by the shape and the color, respectively.

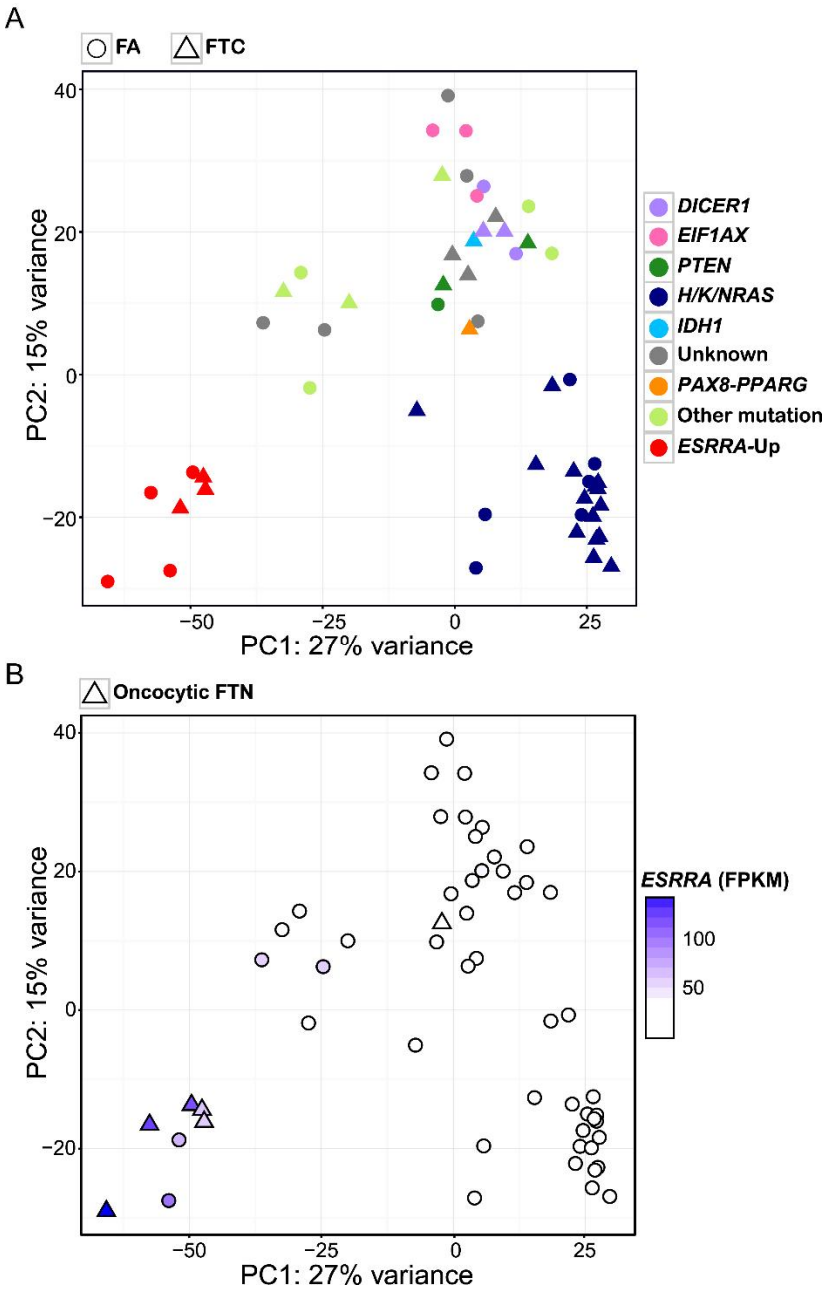

Figure 1-14. The gene expression analysis on miFTC and FA [1]. A) The result of PCA on miFTC and FA. The histological subtype and the driver mutation of each sample were represented by the shape and the color, respectively. B) *ESRRA* up-regulation in oncotytic FTN.

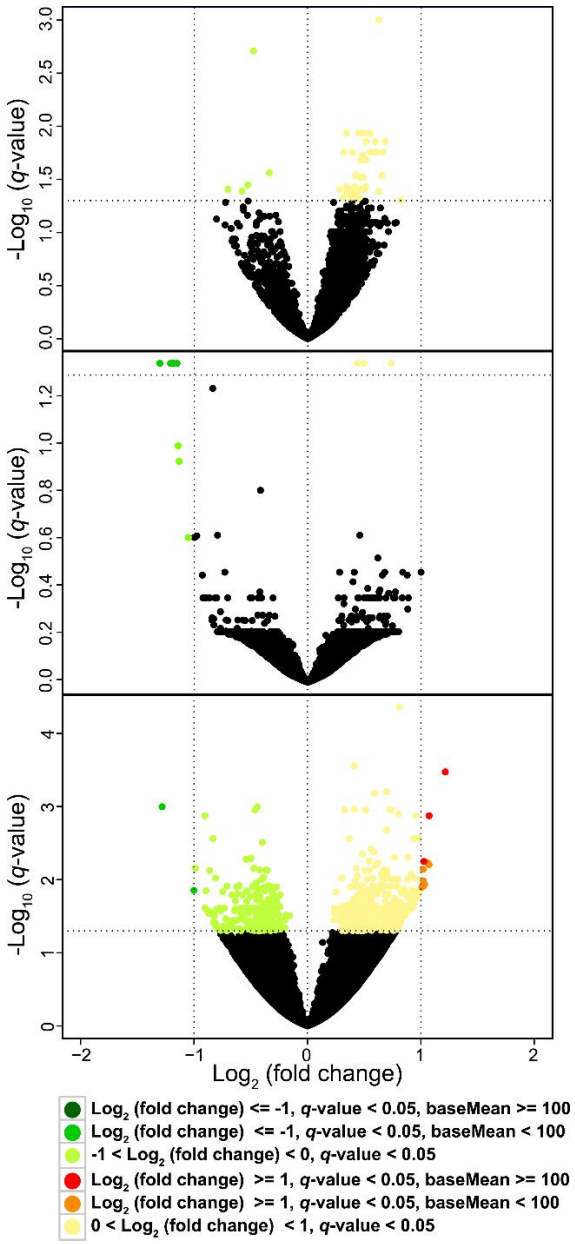

Figure 1-15. DEG analysis on follicular-patterned thyroid tumors [1]. The volcano plots represent identical gene expression among three follicular-patterned thyroid tumors. The comparison of miFTC and FA (upper), FA and EFVPTC (middle), and miFTC and EFVPTC (lower) were shown. All analysis was restricted to *RAS*-like tumors.

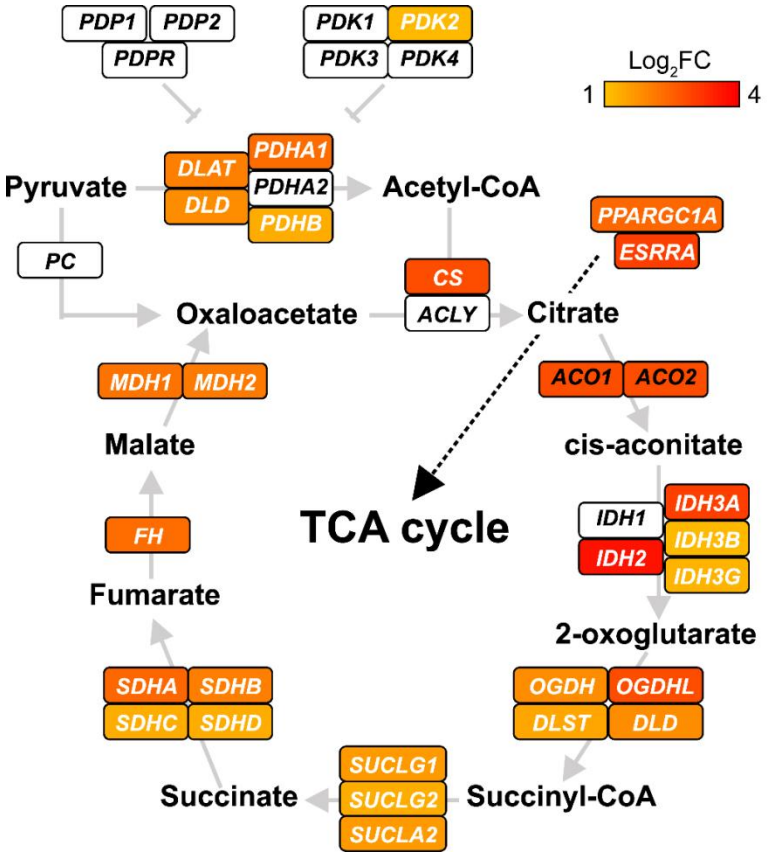

Figure 1-16. The characteristic gene expression of oncocytic FTN [1]. *ESRRA*, *PPARGC1A*, genes in TCA cycle were represented by Log_2FC. This figure was illustrated based on a figure from Davis et al and KEGG pathway (hsa00020) [139, 143].

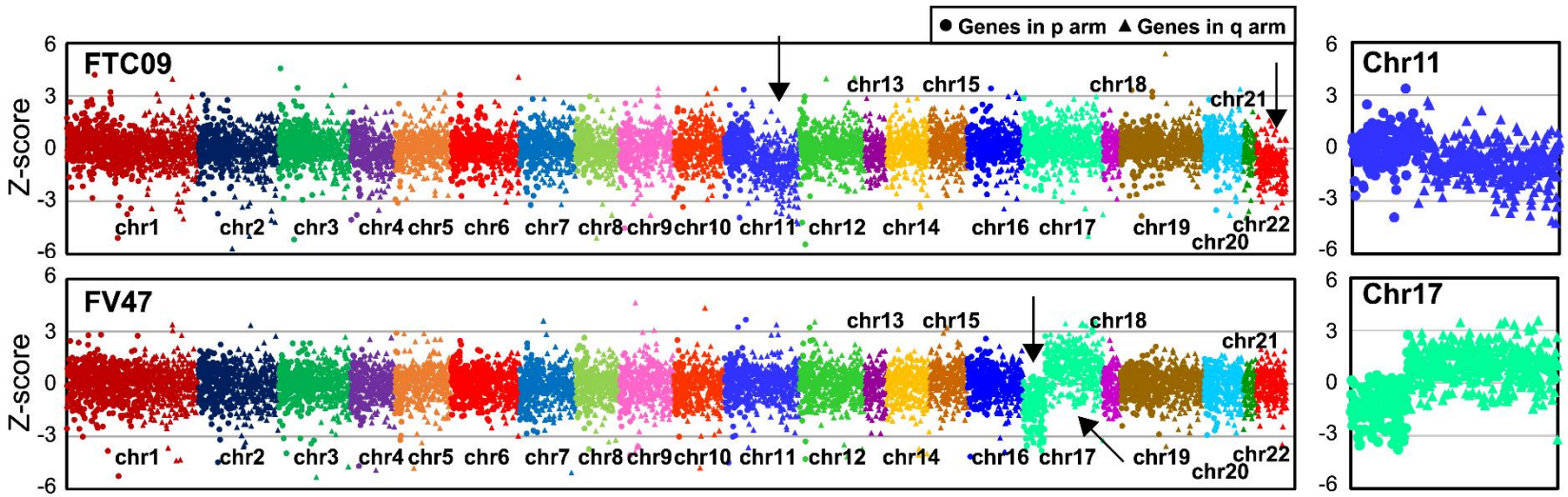

Figure 1-17. The arm-level CNA identification by JRB analysis [1]. The examples of arm-level duplication and deletion were pointed by arrow. The enlarged images of chromosome 11q deletion (upper right) and chromosome 17 with deletion and duplication in p and q arm (lower right) were provided.

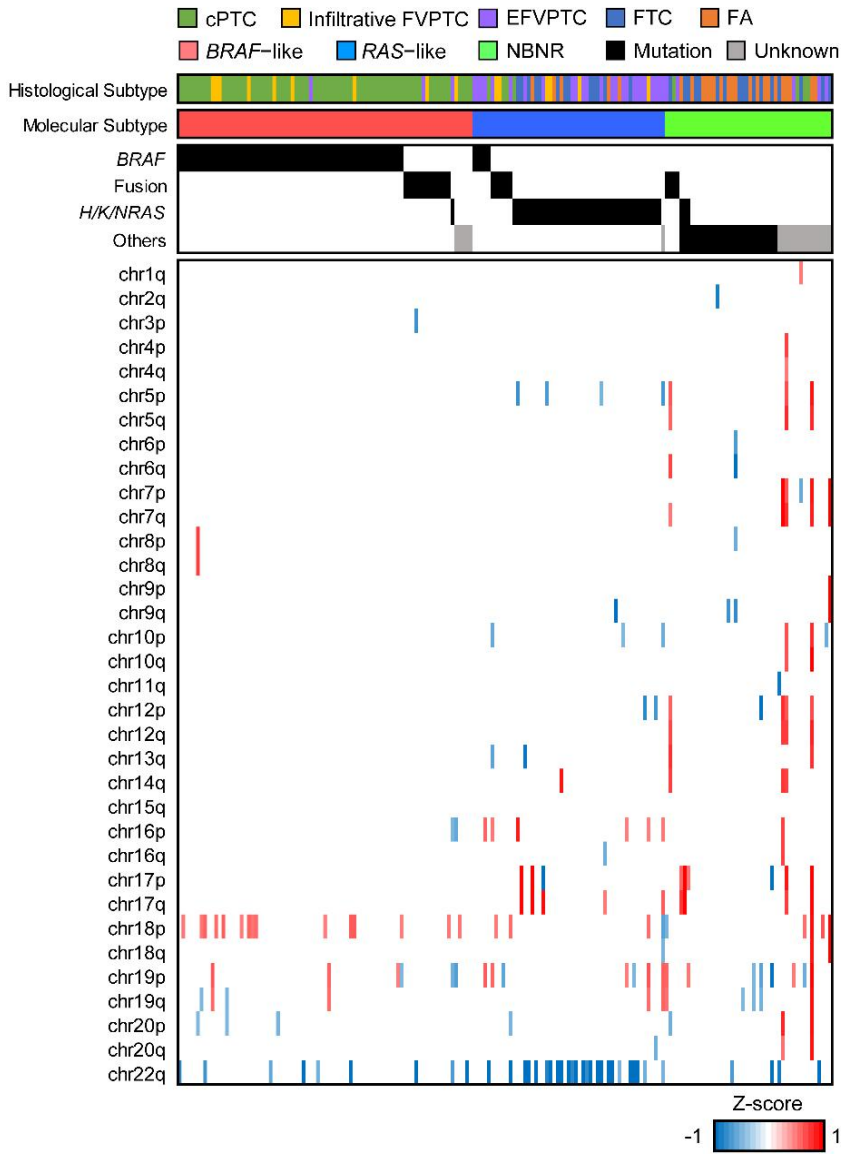

Figure 1-18. The distribution of arm-level CNA across 180 thyroid tumors [1]. The chromosome arms without arm-level CNA across all samples were not shown. Samples were sorted in the same manner as Figure 1-11.

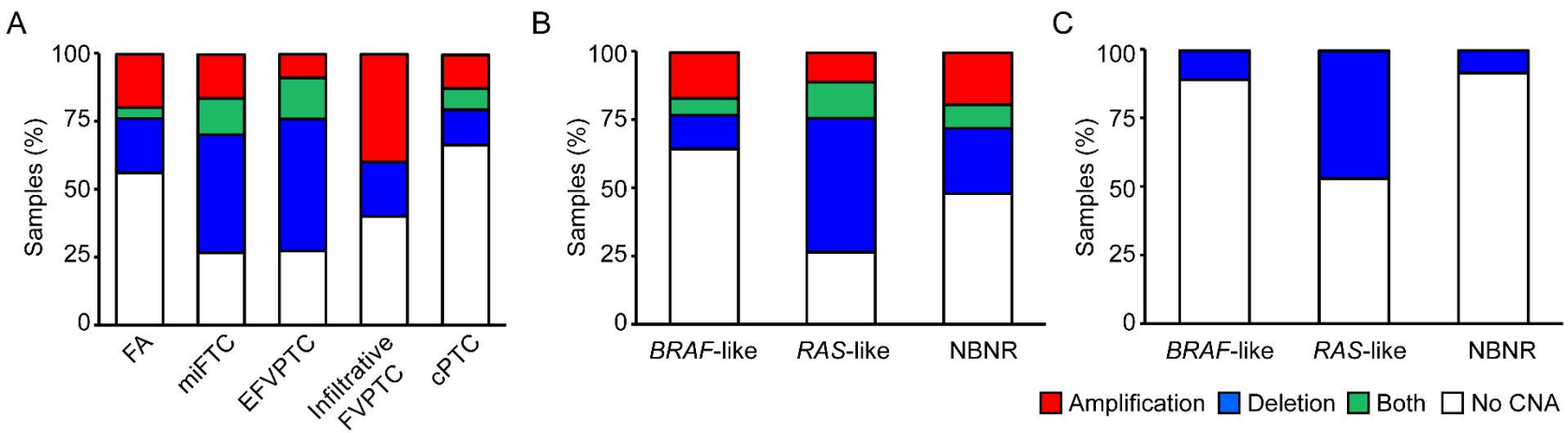

Figure 1-19. The proportion of arm-level CNA in thyroid tumors [1]. The bar charts displayed the proportion of samples with at least one arm-level CNA in thyroid tumors according to A) histological subtype and B) molecular subtype. Amplification, Deletion, and Both denote the sample with only arm-level amplification, only arm-level deletion, and both types of CNA, respectively. C) The proportion of samples with 22q deletion in three molecular subtypes.

Discussion

The molecular characteristics of PTC, the most common type of thyroid cancer, are well established by TCGA [39]. *EIF1AX*, *PPM1D*, *CHEK2*, and diverse fusion genes were newly proposed as oncogenic drivers of this cancer. Meanwhile, FTC which displays a frequent hematogenous spread to lung or bone is not extensively investigated yet [84]. Therefore, the molecular landscape of thyroid cancer, especially WDTC, is not fully revealed.

The well-known driver or candidate genes were found in 72.73%, 89.58%, and 92.21% of FTN, FVPTC, and cPTC, respectively (Figure 1-1 and 1-20). cPTC and FTN displayed a huge difference in the patterns of genetic alteration. As histological features, FVPTC showed an intermediate mutational characteristic between cPTC and miFTC based on their subtypes; EFVPTC and infiltrative FVPTC showed similar mutational profiles to those of miFTC and cPTC, respectively. In consistent with previous studies, the arm-level CNA was more frequently identified in follicular-patterned thyroid carcinoma (miFTC and FVPTC) [39, 144]. To sum up, the aforementioned results support the relationship between type of genetic alterations and histology of thyroid cancers. Also, the lower proportion of arm-level CNA in FA than miFTC indicates that FA is pre-neoplastic form of miFTC although they represented similar mutational profiles.

BRAF-like and *RAS*-like were recently proposed as two molecular subtypes which could be more efficiently applicable to classifying thyroid cancer [39]. However, in this study, three molecular subtypes, *BRAF*-like, *RAS*-like, and NBNR were identified based on transcriptome profiling (Figure 1-9). The result regarding PTC was considerably similar to TCGA study, which classified subtypes of PTC as $BRAF^{V600E}$-like and *RAS*-like. It was reported that FVFTCs in TCGA, which are classified as *RAS*-like, were often misdiagnosed as FTC by pathologists [107]. Moreover, as mentioned earlier, arm-level copy number alterations were frequently observed in FVPTC similarly to FTC as well as *H/K/NRAS* mutations [144]. The aforementioned issues raised a question regarding the distinction between FTC and FVPTC in the point of biological and clinical relevance. In this study, EFVPTC and infiltrative FVPTC showed different mutational and transcriptional characteristics to each other (Figure 1-1 and 1-13). EFVPTC, which was recently re-classified as "noninvasive follicular thyroid neoplasm with papillary-like nuclear features" according to its indolent features [137], had highly similar gene expression profiles to miFTC or FA (Figure 1-15). This result emphasizes that re-classification of thyroid cancers based on their mutational and transcriptional characteristics may be beneficial for stratified medicine.

Discovering potential molecular marker, which could be implemented to differential diagnosis of miFTC and FA was the major objective of the present research. Distinguishing miFTC and FA is important to avoid unnecessary

surgery, but at present-day it is impossible to do so before surgery owing to its highly similar cytological features [84, 106]. There were some suggested molecular markers although they are not practically used. In the present analysis, there was no transcriptional difference between miFTC and FA (Figure 1-14A and 1-15). Moreover, FVPTC which is recently termed as noninvasive follicular thyroid neoplasm with papillary-like nuclear also showed almost same transcriptome profile to them. Aforementioned results suggest a minimal treatment to miFTC and this type of tumor is a very low risk neoplasm.

In TCGA study, *EIF1AX* mutations and *PAX8*-*PPARG* were seemed to contribute the development of *RAS*-like tumor [39]. Meanwhile, the analysis in the present study classified those tumors as NBNR rather than *RAS*-like (Figure 1-9). Thyroid cancer is traditionally defined as the cancer which is developed by the MAPK signaling pathway activation [136, 145], but thyroid cancers with NBNR did not associated with the activation of this signaling pathway (Table 1-7-1-10). Also, these types of tumors displayed distinctive signaling activation among them according to driver gene alteration. Thus, it is expected that more accumulated data with a larger number of tumors with each type of alteration would lead to discover more specific and efficient molecular makers to classify or treatment them. In *BRAF*-like, higher activation of ECM receptor interaction, (CAMs), p53, and MAPK signaling pathways than other molecular subtypes were identified (Table 1-5). Furthermore, low level of TDS and downregulation of several metabolism

pathways supported poor clinical presentation in *BRAF*-like (Figure 1-11 and 1-12). Although there was no establishment of the association between the molecular subtypes and locoregional recurrence (n=1), distant metastasis (n=4), or cancer-specific mortality (n=0), it was clear that *BRAF*-like was strongly related to the other aggressive pathologic characteristics LNM and ETE (Figure 1-11).

Based on the finding of mutational and transcriptional profiles and their clinical association, a schematic model of thyroid cancer progression was illustrated (Figure 1-21).

It is not clear how the oncocytic FTN accumulates an abnormal copy number of mitochondria. In this study, the oncocytic FTNs represented the unique transcriptome profile relative to other FTNs including strong up-regulation of mitochondria-related metabolic pathways (Figure 1-14 and 1-16). The similar consequence was reported in recent study regarding eosinophilic chromophobe renal cell carcinoma which is one type of oncocytoma [143]. *ESRRA* and *PPARGC1A* are reported as stimulating the mitochondrial biogenesis and OXPHOS [146, 147]. It was identified that oncocytic FTNs in the present study also displayed up-regulation of these genes (Figure 1-16). Because this study does not include sequencing data from WGS or WES, the driving alteration of this type of tumor was not fully revealed. However, these findings would lead to an understanding of the association between altered mitochondrial biogenesis and oncocytoma.

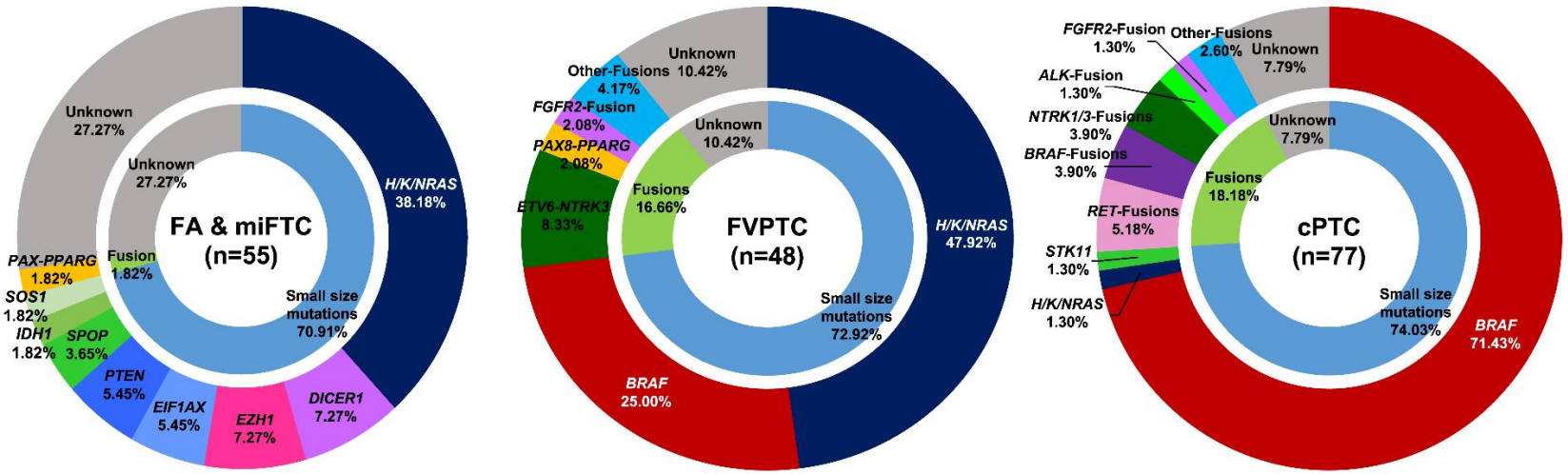

Figure 1-20. The driver mutations in thyroid tumors [1]. Each pie chart represents the distribution of driver mutations identified in FA/miFTC, FVPTC, and cPTC, respectively

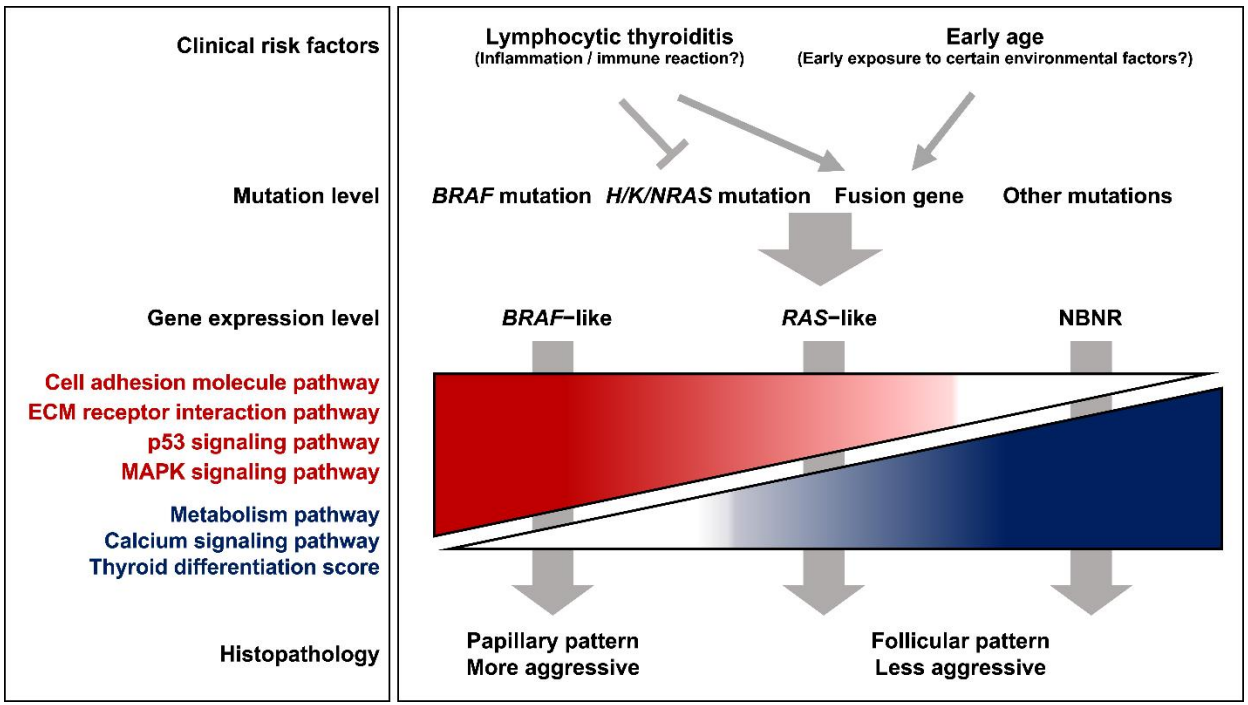

Figure 1-21. A schematic model of thyroid cancer progression [1].

CHAPTER 2.

Integrated analysis of whole-genome, transcriptome, and epigenome in 16 anaplastic thyroid carcinomas

Abstract

Anaplastic thyroid carcinoma (ATC) is a very rare malignancy and one of the most lethal type of cancers that accounts for 50% of deaths by thyroid cancer. Here, whole-genome sequencing with *HpaII* digestion and RNA-sequencing were performed on 16 ATC samples. Moreover, integrative analysis on additional 324 well-differentiated thyroid carcinoma (WDTC) samples, from publicly available or in-house data, demonstrated the genomic, transcriptomic, and epigenomic changes during the progression of ATC from WDTC. As driver genes, *NFE2L2* (6.25%) as well as the most well-known thyroid cancer driver genes, *BRAF* (43.75%) and *H*/*NRAS* (31.25%) were discovered. Also, *TERT* promoter mutations (C228T or C250T) were also recurrently identified (75%). Interestingly, 10 ATC samples had extra somatic mutation or deletion in oncogenes (*EIF1AX*, *AKT1*, and *PIK3CA*), tumor suppressors (*TP53*, *CDKN2A*/*CDKN2B*, *BRCA1*, *PTEN*, *TET1*, and *LATS1*/*LATS2*), and cancer-related genes (*ATRX*, *KMT2C*, and *U2AF1*). In particular, all *RAS*-positive ATC cases had co-mutation in *EIF1AX* and *CDKN2A*/*CDKN2B* loss was identified in 60% of them. Mutation burden and arm-level somatic copy number alterations were significantly increased in ATC ($p < 0.001$ for each) which mirror aggressive genomic features. AID/APOBEC signature in one ATC was discovered from mutational signature analysis and *APOBEC3A* ($Log_2FC = 2.27$) and *APOBEC3B* ($Log_2FC = 1.90$) were overexpressed in ATC relative to WDTC. Interestingly, patients with ATC harbored *TP53*

germline mutations (R49H and E11Q) more frequently than WDTC patients (p = 0.002). Through gene expression analysis, a cluster which distinguishes ATC from WDTC was determined and proposed as a novel molecular subtype, ATC-like. ATC showed an extreme loss of differentiation and the activation of MAPK signaling pathway. With weighted co-expressed gene analysis, in *RAS*-positive ATC, a network (R = 0.99 and $p < 0.001$) with up-regulated *IL6* ($Log_2FC = 6.49$) which triggers JAK/STAT signaling pathway and cell cycle was discovered. Notably, the elevated expression of *CD274* (PD-L1; $Log_2FC = 4.17$) and *PDCD1LG2* (PD-L2; $Log_2FC = 4.39$) were observed as well as *IL6* which can be considered as potential immunotherapeutic targets. In particular, the worst thyroid cell differentiation and up-regulation of *CD274* and *PDCD1LG2* was tightly associated with *CDKN2A*/*CDKN2B* deep deletion. From DNA methylation analysis, the global DNA methylation pattern of ATC was distinguished from WDTC and the more frequent hypermethylation of ATC, enriched in CpG island ($p < 0.001$) than WDTC, was verified. Taken together, conclusions could be made that the progression of ATC is delivered by additional hits of somatic mutation in *TERT*, oncogenes, tumor suppressors, or regulatory genes on WDTC. Furthermore, ATC suffers several extreme transcriptional and epigenomic alterations including JAK/STAT and MAPK signaling pathways activation and CpG island hypermethylation. This study will broaden the understanding of this lethal disease and allow more efficient diagnostic strategy and attractive targeted therapeutic approach in future.

Keywords: Thyroid carcinoma; Anaplastic thyroid carcinoma; Massively-parallel sequencing; Whole-genome sequencing; RNA sequencing

Student number: 2014-21325

Introduction

ATC or undifferentiated thyroid carcinoma is one of the most aggressive human cancers and represents up to 50% of deaths by thyroid carcinoma [86, 87]. ATC is a very rare and only accounts for less than 2% of thyroid carcinoma [85]. It is usually develops from preceding lesion of WDTC or PDTC, but also arise *de novo* [89, 90]. The patients with ATC present with a rapidly enlarging anterior neck mass, dyspnea, dysphagia, and voice change [93, 94]. Because of its aggressive features, early diagnosis of this malignancy is important for the patient survival. However, the majority of patients with ATC are diagnosed at advanced stage or have distant metastasis to other organs which leads to poor prognosis with conventional therapies in the present [89].

However, besides the clinical implication of ATC, there is insufficient number of large scale studies on ATC using MPS technology. Aforementioned studies well described genomic and transcriptomic features of ATC. However, their analyses have weak points such as insufficient number of the patient-derived ATC tumors, the low resolution of method, and the absence of epigenomic analysis. Also, they were focused on PDTC or ATC without accompanying WDTC. Thus, the molecular pathogenesis and progression of ATC from WDTC, including PTC and FTC, are not fully elucidated yet.

In this study, WGS with *HpaII* digestion and RNA-seq was performed on 16 ATC and 15 matched normal samples (13 blood and 2 thyroid tissue samples) to reveal the genomic, transcriptomic, and epigenomic alterations during progression of ATC from WDTC. With high sequencing coverage and in-depth analysis, I focused on identifying novel genes which encourage aggressiveness of thyroid carcinoma. It is expected to discover key pathways or gene regulations which promote WDTC to ATC integrating with data of 324 WDTC samples from publicly available or in-house.

Materials and Methods

Patients

The clinicopahtological characteristics of 16 study subjects were described in Table 2-1. The specialized pathologist reviewed the pathological slides. The patient was diagnosed with ATC when the whole tumor was ATC or partial WDTC component was identified but the portion was less than 10%. There were three focal PD/ATC (ATC1-T, ATC2-T, and ATC4-T) with background of PTC.

Whole-genome sequencing pipeline

Prior to sequencing library preparation, DNA fragmentation was performed by *HpaII* restriction enzyme. Then, WGS was performed using Hiseq X instrument based on manufacturer's instruction (Illumina, SA, San Diego). The raw sequenced 150 bp paired-end reads were aligned to hg19 reference genome using BWA algorithm [148]. PCR duplicate reads were removed by Picard tools (https://broadinstitute.github.io/picard/). Indel realignment and BQSR were conducted by GATK based on best practice (https://software.broadinstitute.org/gatk/best-practices/) [113].

RNA sequencing pipeline

RNA-seq was performed using Hiseq 2000 instrument based on manufacturer's instruction (Illumina, SA, San Diego). The raw sequenced 101

bp paired-end reads were aligned to hg19 reference genome using STAR-2pass algorithm [112]. Duplicated read removal was achieved by Picard tools (https://broadinstitute.github.io/picard/). Split N cigar reads, indel realignment and BQSR were applied to RNA-seq data based on GATK's best practice (https://software.broadinstitute.org/gatk/best-practices/) [113].

Whole-exome sequencing pipeline

To compare genomic characteristics of ATC, WES data of WDTC from previously published studies were processed. There were 18 FTNs including five FA, eleven miFTC, and two widely invasive FTC (wiFTC) samples from study of Jung et al [83] and In-house nine wiFTC samples were also used. Moreover, 28 PTC samples from study of Rubinstein et al was analyzed [149]. The sequenced read files were downloaded from European Nucleotide Archive (https://www.ebi.ac.uk/ena). If ready-to-download sequenced read files were not available, sra files were downloaded from Short Read Archive (https://www.ncbi.nlm.nih.gov/sra) and decompressed by SRA Toolkit [150]. Also, matched normal samples for each tumor were managed. Aforementioned whole-exome sequenced samples were processed by same pipeline for WGS.

Identification of somatic SNV and indel

The somatic SNVs and indels were discovered using MuTect and GATK's HaplotypeCaller [44, 151]. For ATC samples, these methods were applied to

WGS and RNA-seq data. The somatic mutations which were found in at least one sequencing method were used for the subsequent analysis. The low complexity regions defined by 1000 genomes project were not processed [152]. For whole-exome sequenced samples, MuTect was used to call somatic SNVs. Discovered variants were annotated by ANNOVAR [114]. The variants under MAF of 0.1% based on Exome Aggregation Consortium (version 0.3 without TCGA) were remained for subsequent analysis [115].

Mutational signature analysis

SomaticSignatures package was used for determine the mutational signatures of 15 ATC tumors which have matched normal sample [153]. All the passed SNV calls from MuTect algorithm was subjected to the analysis [44]. The reference mutational signatures were downloaded from COSMIC (http://cancer.sanger.ac.uk/cosmic/download). The mutational signature assignment of each ATC sample was performed using average linkage hierarchical clustering algorithm with 30 reference signatures.

Identification of somatic copy number alteration

To identify the SCNAs in ATC, FACETS was implemented to WGS data [154]. To prepare the input data, the sequencing depth of germline variants based on dbSNP142 was calculated by mpileup of samtools [155]. For WES data, EXCAVATOR2 was used [64]. Furthermore, GISTIC2 detected the significantly altered genes, regions, and chromosome arms of each subtypes

[156]. Arm-level SCNA was determined when each chromosome arm bears the total size of duplication or deletion over than 50% of chromosome itself.

Differentially expressed gene analysis

The individual gene expression level was quantified using HTSeq [117]. Then, DESeq2 was used to call DEGs [21]. In this study, DEGs were determined based on following criteria; i) $Log_2FC >= |1|$ and ii) adjusted *p*-value < 0.05. For this analysis, three ATC samples which are diagnosed as focal PD/ATC (ATC1-T, ATC2-T, and ATC-T) and ATC12-T which represents similar gene expression pattern to focal PD/ATC as PCA suggested were excluded.

Weighted gene co-expression analysis

Weighted gene co-expression analysis (WGCNA) method was applied to reveal the central gene network which promotes the progression of ATC from WDTC [157]. The normalized expression values from variance stabilizing transformation (VST) method from DESeq2 were used for WCGNA [21]. The analysis was performed twice to establish the key networks which stimulates i) FTC origin ATC (using *RAS*-positive samples) and ii) PTC origin ATC (using *BRAF*-positive samples). WGCNA method uses the first principal component of genes in each network to determine the association between the networks and the clinical information (histology of samples in this study). The significance of their relationships were verified by Pearson's correlation coefficient and *p*-value.

Pathway enrichment analysis

The pathway enrichment analysis using KEGG database was conducted by DAVID [139, 158].

Thyroid differentiation score analysis

To investigate the differentiation of ATC tumors, TDS which was recently proposed by TCGA study was calculated [39]. Using 16 genes which are related to thyroid metabolism such as thyroid hormone signaling, thyroid morphogenesis, and thyroid cell differentiation, *DIO1*, *DIO2*, *DUOX1*, *DUOX2*, *FOXE1*, *GLIS3*, *NKX2-1*, *PAX8*, *SLC26A4*, *SLC5A5*, *SLC5A8*, *TG*, *THRA*, *THRB*, *TSHR*, and *TPO*, TDS was calculated as the average value of median-subtracted normalized expression across 16 genes. In concordance to WGCNA, the VST expression values from DESeq2 were used [21].

ERK score analysis

To determine the activation of ERK/MAPK signaling pathway of ATC, the modified TCGA's ERK scoring method from the first chapter of this thesis was implemented [39]. Using 52 genes involved in MAPK signaling pathway [121], ERK score was calculated as the average value of median-subtracted normalized expression across 52 genes. In concordance to WGCNA and TDS analysis, the VST expression values from DESeq2 was used [21].

DNA methylation analysis

Methylation sensitive restriction enzyme, *HpaII*, recognizes and cuts CCGG sites, if second cytosine is unmethylated [159]. Using this characteristic of *HpaII*, the global DNA methylation pattern could be investigated with WGS (Figure 2-1A). After alignment of paired-end reads which were sequenced using *HpaII* digested DNA fragment to GRCh37 reference genome, the mapped reads which span 2,297,198 CCGG sites of hg19 were extracted using BEDTools [160]. Then, the DNA methylation level of second cytosine was calculated based on 10 possible patterns of mapped read with mapping quality over 20 (Figure 2-1B). Four patterns of mapped reads which spanning CCGG site but do not contain information of DNA methylation status were excluded from calculation. Only sites which covered more than five mapped reads across 16 ATC samples were used for further analysis.

Table 2-1. The clinicopathological characteristics of 16 ATC patients.

ID	Main histology	Sex	Age, years[a]	Tissue analyzed	Preceding WDTC[b]	Sites of distant metastasis	DSS, months[c]	Final status	Site responsible for main problem
ATC1-T	PTC (10% PD)	F	63.6	PT	PTC	None	15	DOD	Primary
ATC2-T	PTC (10% PD)	F	91.4	PT	PTC	Lung	3.7	DOD	Primary
ATC3-T	PD/ATC	M	52.2 [57.5]	LRT (LN)	FTC	Lung, bone	99.7 [35.4]	AWD	Distant
ATC4-T	PTC (5% ATC)	F	(48.9) 75.6	LRT (trachea)	PTC	Lung, bone	18.6	AWD	Primary
ATC5-T	PD/ATC	F	(58.8) 64.2 [65.7]	LRT (LN)	PTC	Lung, bone	33.3 [15.9]	DOD	Primary
ATC6-T	ATC	F	(71.8) 81.9	PT	FTC	Lung	12.5	DOD	Primary/distant
ATC7-T	ATC	F	68.9	PT	unknown	Lung, adrenal gland	1.0	DOD	Primary
ATC8-T	ATC	M	67.8 [68.7]	DMT (soft tissue)	FTC	Brain, soft tissue (buttock), lung	14.7 [3.3]	DOD	Distant
ATC9-T	ATC	F	(44.3) 55.7	DMT (brain)	PTC	Lung, bone, brain	1.9	DOD	Distant
ATC10-T	ATC	F	75.2	PT	PTC	Lung	4.9	DOD	Primary
ATC11-T	ATC	F	54.1	PT	FTC	Lung, soft tissue (psoas)	0.9	DOD	Primary/distant
ATC12-T	ATC	M	68.0	PT	PTC	Lung	3.1	DOD	Distant

Table 2-1. continued.

ID	Main histology	Sex	Age, years[a]	Tissue analyzed	Preceding WDTC[b]	Sites of distant metastasis	DSS, months[c]	Final status	Site responsible for main problem
ATC13-T	ATC	M	71.2	PT	PTC	Lung	11.4	FU loss	Distant
ATC14-T	ATC	F	55.7 [55.8]	LRT (vessel)	FTC	Lung, vessel, esophagus, duodenum	14.2 [12.9]	DOD	Primary/distant
ATC15-T	ATC	M	65.5 [65.6]	LRT (LN)	PTC	Lung, bone	2.6 [1.4]	DOD	Primary
ATC16-T	ATC	F	71.5	PT	PTC	-	8.9	FU loss	NA

Abbreviations: WDTC, well-differentiated thyroid cancer; DSS, disease-specific survival; ATC, anaplastic thyroid cancer; FTC, follicular thyroid cancer; PTC, papillary thyroid cancer; PD, poorly-differentiated; PT, primary tumor; LRT, locally recurred or residual tumor; DMT, distant metastatic tumor; LN, lymph node; DOD, death of disease; AWD, alive with disease; FU, follow-up; NA, non-applicable

[a] age at diagnosis of ATC was shown. If there existed preceding WDTC before the diagnosis of ATC, the age was added as (age at diagnosis of preceding DTC), and if the age at the tissue obtained was different from that of the diagnosis of ATC, the ages was also added as [age at which the analyzed tissue was obtained].

[b] determined by the co-existing WDTC components or mutational status of $BRAF^{V600E}$

[c] defined as the length of time from the date of diagnosis of ATC/ [the operation date of analyzed tissue] to the date of last follow-up including death. If the operation date for the tissue analyzed was same to the data of ATC diagnosis, only one is presented.

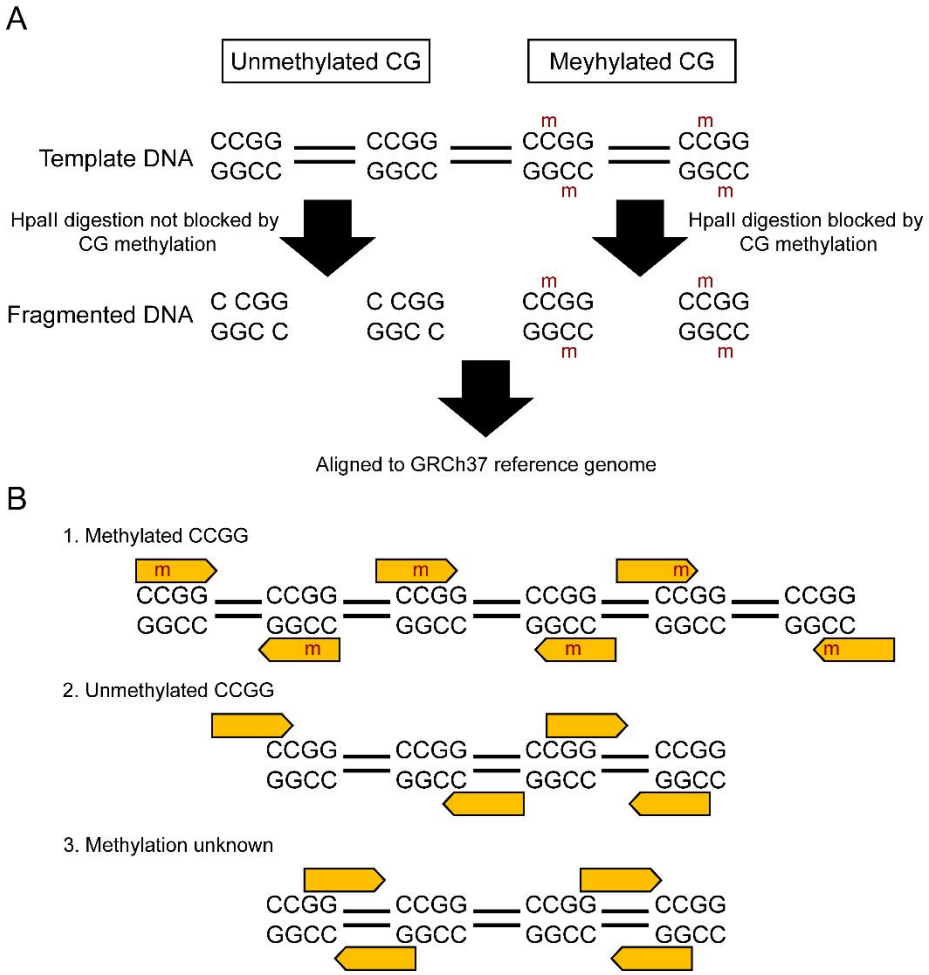

Figure 2-1. The schematic illustration of *HpaII* restriction enzyme treatment to study genome-wide DNA methylation level. A) The effect of *HpaII* digestion according to methylation status of CCGG sites. B) The sixteen possible sequenced read alignment on CCGG sites.

Results

Massively-parallel sequencing summary of study subjects

WGS summary statistics of 16 ATC and 15 matched normal samples were provided in Table 2-2 and Table 2-3, respectively. There were 13 matched normal blood and two normal thyroid tissue samples (ATC15-N and ATC16-N). Meanwhile, ATC12-T does not have a matched normal sample. Tumor and normal samples were sequenced by the average sequencing depth of 74.75 X and 36.85 X, respectively. In case of RNA-seq, 16 ATC tumors were sequenced by 6.39 Gb on average (Table 2-4).

Mutation profile of anaplastic thyroid carcinoma

BRAF^{V600E} mutation was the most frequently identified driver mutation in ATC with 43.75% of incidence (Figure 2-2). It is well reported that *BRAF*^{V600E} is the most recurrently identified somatic driver mutation in PTC with 53~59% incidence [39]. *RAS* (*HRAS* and *NRAS*; 31.25%) was the second most frequently mutated driver gene in ATC; *HRAS*^{G13R} (ATC8-T), *HRAS*^{Q61R} (ATC7-T), and *NRAS*^{Q61R} (ATC11-T, ATC13-T, and ATC14-T) were identified. Mutations in *HRAS, NRAS*, *KRAS* were frequently detected in FA, miFTC, and encapsulated FVPTC in the first chapter of this thesis. The aforementioned results suggested that study subjects in this study arose from both PTC and FTC.

Four samples (ATC3-T, ATC6-T, ATC10-T, and ATC15-T) did not have well-known thyroid cancer driver gene mutation, but all of them had mutation in cancer-related genes. ATC3-T harbored *NFE2L2*^{G65A}. *NFE2L2* is reported as cancer driver gene which interacts with *KEAP1* and *CUL3* and 15% of lung squamous cell carcinoma harbors mutation in this gene [161, 162]. Recently, *PAX8-NFE2L2* fusion gene driver was identified in thyroid cancer [31]. ATC10-T had *TP53*^{R141H} which is the most frequently altered TSG in human cancers [163]. In addition, ATC6-T and ATC15-T had *TET1*^{T1150fs} and *KMT2C*^{Y987X} which are frequently mutated in several types of cancer [164-166].

Furthermore, 75% of ATC samples had *TERT* promoter mutations; C228T (nine out of 16 samples) and C250T (one out of 16 samples; Figure 2-2 and 2-3). Only ATC5-T harbored both C228T and C250T. In ATC4-T, a somatic mutation in *ATRX*^{S343X}, a chromatin remodeler, was found which represses alternative lengthening of telomere (Figure 2-2) [167, 168].

According to previous two large scale studies regarding thyroid cancer, it is now well known that the alterations of cancer genes in WDTC are highly mutually exclusive; most tumors harbor mutation in single driver gene such as *BRAF*, *HRAS*, *NRAS*, *EIF1AX*, and *DICER1*, but there is no additional mutation in other oncogene or TSG [39]. However, 75% of study subjects had additional mutations in cancer-related genes in this study. Five samples with *RAS* mutations had two positions of splicing mutation (338-1G>T and 338-

2A>G in ATC7-T and ATC8-T, respectively) and one type of missense mutation (G8E in ATC11-T, ATC13-T, and ATC14-T) in *EIF1AX* which is recently reported as thyroid cancer driver gene [39, 124].

Moreover, somatic and germline mutations of *TP53* were discovered in five study subjects. Four samples (ATC5-T, ATC7-T, ATC8-T, and ATC16-T) harbored somatic mutations in *TP53* and two samples (ATC7-T and ATC13-T) had germline *TP53* mutations (R49H and E11Q; Table 2-5). To determine the potential association between *TP53* germline mutation and ATC, *TP53* germline mutation status of WDTC patients of previous studies or in-hose data was also determined [39, 83, 149]. The patients with ATC more significantly harbored germline mutation in *TP53* than WDTC (Fisher's exact test $p < 0.001$; Table 2-5) and there was only one WDTC patient with *TP53* germline mutation (E11Q). Interestingly, this patient was diagnosed with wiFTC with bone metastasis; wiFTC is aggressive type of WDTC and often represetants aggressive features such as hematogenous spread and distant metastasis [169].

Two samples (ATC8-T and ATC12-T) had $AKT1^{E17K}$ mutation. *AKT1* is well known oncogene and recurrently identified in various cancers and most somatic mutation in *AKT1* is E17K [170-172]. ATC16-T had $PIK3CA^{H1047R}$ which is well known oncogenic mutation [173, 174]. Recently, one study reported that $PIK3CA^{H1047R}$ mutation is co-occurred in 10.3% of ATC tumor with $BRAF^{V600E}$ but not in WDTC [175]. $PTEN^{T319fs}$ was found in ATC3-T

which also has *NFE2L2*^{G65A} mutation and ATC9-T had *LATS1*^{R697fs} and *LATS2*^{R961fs} which are reported as TSG [176, 177]. Furthermore, somatic hotspot mutation, *U2AF1*^{S34F}, which is spliceosomal machinery gene was identified in ATC2-T. It is frequently identified in lung adenocarcinoma and known to be involved in progression of acute myeloid leukemia [178-180]. The detailed annotation of each mutation is provided in Table 2-6.

As previously described, there were several additional hits of somatic mutations in ATC. Fortunately, the further analysis on our in-house data for wiFTC which is one of the aggressive type of WDTC instructed the impact of additional mutations to thyroid carcinoma progression. All wiFTC samples with *H/NRAS* mutations (n=7) had *TERT* promoter mutation or SV (Figure 2-4). However, there was only one sample with additional altered driver gene (*NRAS* and *EIF1AX* co-mutation). This result emphasized that the order of gene alteration during thyroid carcinoma progression. It corresponds to the alterations in genes with following orders; i) driver genes (e.g. *BRAF*, *H*/*K*/*NRAS*, or *EIF1AX*), ii) *TERT* (promoter mutation or SV), and iii) oncogene (e.g. *AKT1*, *EIF1AX*, *or PIK3CA*), TSG (e.g. TP53, *BRCA1*, or *PTEN*), or regulatory genes (e.g. *ATRX* or *U2AF1*).

Mutational signature analysis

The mutational signature of ATC which suggests the etiology of cancer was determined using 15 ATC samples [41]. The hierarchical clustering analysis

was performed using 15 mutational signatures from study subjects and 30 reference signatures which were provided by COSMIC. The signature 5 was found in 14 out of 15 samples which is the most frequently identified signature in various types of cancer including PTC and FTC though their etiology is not revealed (Figure 2-5A). Interestingly, ATC4-T represented signature 2 which is derived by the activation of AID/APOBEC family of cytidine deaminases (Figure 2-5B) [181, 182]. In particular, the expression levels of *APOBEC3B* (Log_2FC = 1.90) and *APOBEC3A* (Log_2FC = 2.27), which are the major and second contributors of this signature, were up-regulated in ATC relative to WDTC (Figure 2-6) [183-185].

Increased tumor mutation burden in anaplastic thyroid carcinoma

It is well-known that mutation burden of cancer genome is increased during advanced or progression which reflects their aggressiveness [82, 186-188]. As expected, tumor mutation burden of ATC (3.50 mutations/Mb on average) was elevated compared with FTN (1.53 mutations/Mb on average) and PTC (1.33 mutations/Mb on average; Mann-Whitney test $p < 0.001$; Figure 2-7A).

Somatic copy number alteration in anaplastic thyroid carcinoma

The SCNA in ATC tumors was determined with two different resolutions; i) arm-level SCNA and ii) focal SCNA. There were only two significantly altered chromosome arms including amplified 21p ($q < 0.001$) and 21q (q = 0.02; Table 2-7). In terms of the focal SCNA, there were 12 and eight regions

with significant amplification (2q21.2, 16q22.2, 6p11.2, 15q26.3, 1p36.13, 1q21.1, 7q11.22, 17q12, 2p11.1, 14q11.2, 10q11.22, and 15q26.3) and deletion (9p21.3, 2q36.1, 19p13.3, 8p23.1, 22q13.32, 9q22.2, 17p11.2, and 21q22.3), respectively (Table 2-8 and 2-9).

From the above-mentioned regions, most of them did not cover the cancer-related genes. However, 9p21.3 contained *CDKN2A*/*CDKN2B* and four samples (ATC8-T, ATC9-T, ATC11-T, and ATC14-T) had loss of these genes (Figure 2-8A). *CDKN2A*/*CDKN2B* are TSGs and their deletion is frequently discovered in glioblastoma and nasopharynx cancer [35, 56]. Interestingly, three out of four tumors had *RAS*/*EIF1AX* co-mutation and one other tumor (ATC9-T) had $BRAF^{V600E}$ mutation (Figure 2-8A). This consequence was not reported in previous thyroid cancer studies (Figure 2-8B) [39]. Only single case with $BRAF^{V600E}$ and *CDKN2A*/*CDKN2B* deletion was reported in recent study regarding PDTC/ATC [101]. Also, this type of deletion was screened in thyroid cancer cell lines from Cancer Cell Line Encyclopedia (CCLE) [189], and two ATC cell lines (CAL62 and SW579) among thyroid follicular epithelial cell origin cancers represented loss of *CDKN2A*/*CDKN2B* (Table 2-10).

In some types of cancer, SCNA burden is correlated to patient survival or cancer aggressiveness [190-192]. To verify the relationship between SCNA and thyroid cancer aggressiveness, the frequency of arm-level SCNA in ATC was compared with WDTC. As the mutation burden was increased in ATC,

the arm-level SCNA burden was also abundant in ATC (Mann-Whitney test $p < 0.001$; Figure 2-7B). Because it is well known that PTC, usually *BRAF*-positive, display low SCNA burden relative to FTC, usually *RAS*-positive, the arm-level SCNA burden of ATC was again compared with WDTC as two groups, *BRAF*-positive and *RAS*-positive, respectively. *BRAF*-positive ATC displayed the significant increment of arm-level SCNA relative to *BRAF*-positive WDTC (Mann-Whitney test $p < 0.001$; Figure 2-7C). *RAS*-positive ATC displayed higher but similar arm-level SCNA burden compared with FA/miFTC and wiFTC, respectively (Mann-Whitney test $p = 0.01$ and $p = 0.59$, respectively.).

Molecular subtypes of anaplastic thyroid carcinoma

To classify the molecular subtype of 16 ATC tumors based on their transcriptome profile, K-means clustering via PCA was performed incorporating data from 170 WDTC samples from the first chapter of this thesis. To determine the most appropriate classification, K-means clustering was conducted from K=1 to K=10 and the optimal K was determined as four according to elbow method (Figure 2-9). The majority of ATC samples were separated from three molecular subtypes of WDTC (*BRAF*-like, *RAS*-like, and NBNR; Figure 2-10) and belonged to fourth cluster. This molecular subtype, ATC-like, was composed of samples with diverse types of driver genes including *BRAF*, *H/NRAS*, and *NFE2L2*. Based on this finding, it could be considered that the initial molecular features of WDTC may differ according

to the driver mutation such as *BRAF* and *RAS,* but they tend to represent similar transcriptomic features regardless of the driver mutations ate the end of advancement. Notably, four samples with *BRAF* mutation were categorized with *BRAF*-like and three of them was focal PD/ATC. Lastly, one sample with *TET1* frameshift mutation was classified as NBNR.

Differentiation of anaplastic thyroid carcinoma

Next, the thyroid cell differentiation of ATC was investigated by TDS analysis from TCGA [39]. It was found that ATC shows the significant decrement of TDS (Mann-Whitney test $p < 0.001$ and $p = 0.0011$ for compared with FTN and PTC, respectively; Figure 2-11A). More particularly, *RAS*-positive ATC samples represented the lowest TDS relative to not only the WDTC, but also *BRAF*-positive ATC samples (Mann-Whitney test $p = 0.0003$). It was very drastic loss of differentiation, since *RAS*-positive WDTC shows good thyroid cell differentiation. Among 16 genes, *RAS*-positive ATC displayed significant down-regulation in 15 genes (except for *THRA*) compared with *RAS*-positive WDTC (Figure 2-12): *TSHR* ($Log_2FC = -10.61$), *NKX2-1* ($Log_2FC = -9.47$), *TG* ($Log_2FC = -9.61$), *DIO2* ($Log_2FC = -7.92$), *TPO* ($Log_2FC = -10.01$), *DUOX2* ($Log_2FC = -8.57$), *SLC26A4* ($Log_2FC = -7.71$), *DUOX1* ($Log_2FC = -4.64$), *FOXE1* ($Log_2FC = -5.80$), *SLC5A8* ($Log_2FC = -8.58$), *DIO1* ($Log_2FC = -7.12$), *GLIS3* ($Log_2FC = -2.11$), *THRB* ($Log_2FC = -1.74$), and *SLC5A5* ($Log_2FC = -3.48$), and *PAX8* ($Log_2FC = -1.57$). In contrast to *RAS*-positive ATC, seven out of 16 TDS genes in *BRAF*-positive ATC were significantly

down-regulated relative to *BRAF*-positive WDTC: *TSHR* (Log_2FC = -2.17), *GLIS3* (Log_2FC = -1.20), *THRB* (Log_2FC = -1.49), *FOXE1* (Log_2FC = -1.22), *TG* (Log_2FC = -1.72), *DUOX2* (Log2FC = -1.52), and *DIO2* (Log_2FC = -1.25).

MAPK pathway activation in anaplastic thyroid carcinoma

The constitutive activation of MAPK signaling pathway in thyroid cancer, especially *BRAF*-like, is well described in previous studies [39, 136]. The activation of MAPK signaling pathway in ATC was measured using modified ERK score approach from the first chapter of this thesis [39]. ATC, especially *RAS*-positive, showed increment of MAPK signaling pathway (Figure 2-11B and 2-12; Mann-Whitney test $p < 0.001$). However, TDS and ERK score of *BRAF*-positive ATC were not statistically higher than those of *BRAF*-positive WDTC (Mann-Whitney test p for 0.17 and 0.08, respectively.). Similar to the first chapter of this thesis, TDS and ERK score represented strong negative correlation (Spearman's rank correlation coefficient R^2 = -0.79 and $p < 0.001$).

Weighted gene co-expression analysis of anaplastic thyroid carcinoma

WGCNA was conducted to identify the key gene network (module) which triggers aggressiveness and progression of ATC. The analyses were performed twice using i) *RAS*-positive FTNs and *RAS*-positive ATCs which were developed from preceding FTC and ii) *BRAF*-positive PTCs and *BRAF*-positive ATCs which are developed from pre-existing PTC.

In case of *RAS*-positive, eight modules were significantly correlated to progression (Figure 2-13). In particular, yellow module showed almost perfect correlation coefficient to development of FTC origin ATC progression (Pearson's correlation coefficient R = 0.99 and $p < 0.001$). Then, pathway enrichment analysis using KEGG pathway was conducted to verify the functional relevance of each module. In yellow module, there were several cancer-related pathways including cell cycle, p53 signaling pathway, and JAK/STAT pathway. Furthermore, thyroid hormone synthesis was also enriched which again confirms its significance in thyroid cancer progression as the decrement of TDS in prior analysis was suggested.

Then, JAK/STAT signaling pathway was visualized to determine which genes in this pathway were dysregulated during progression (Figure 2-14). As upstream signals, the expression level of several inflammatory molecules and their receptors were elevated. Among inflammatory molecules, *IL6* (Log_2FC = 6.49) was up-regulated in *RAS*-positive ATC relative to FTN (Figure 2-15). Through the pathway, *MCL1* (Log_2FC = 1.48) and *MYC* (Log_2FC = 4) which are involved in anti-apoptosis and cell-cycle progression and regulated by *IL6* were overexpressed (Figure 2-15) [193, 194].

In addition to cancer-related pathways, the significant up-regulation of *CD274* which encodes PD-L1 was identified in yellow module (Log_2FC = 3.7; Figure 37A). Moreover, *PDCD1LG2* which encodes PD-L2 was found in black

module which is associated with immune pathways such as cytokine-cytokine interaction and it was overexpressed in ATC ($Log_2FC = 4.12$; Figure 2-16B). Both *CD274* and *PDCD1LG2* are receiving attention as immunotherapeutic targets in cancer treatment [195, 196]. Interestingly, ATC samples with *CDKN2A*/*CDKN2B* loss exhibited up-regulation of these two PD-1 ligands and they also suffered the most drastic loss of differentiation (Chi-square test $p = 0.003$; Figure 2-17A). Moreover, their similarity in transcriptome level was further measured by PCA (Figure 2-17B). The similar consequence was detected in one ATC cell line; deep deletion of *CDKN2A*/*CDKN2B* and *CD274* (PD-L1) up-regulation were both found in CAL62 cell line (Table 2-11).

In case of *BRAF*-positive ATC progression, the gene network was not well constructed as the analysis using *RAS*-positive samples. It might be derived by the severe gene expression variance within *BRAF*-positive ATCs in contrast to *RAS*-positive ATCs (Figure 2-18). Thus, DEG analysis was performed to identify key regulated pathway which triggers *BRAF*-positive ATC progression. Among ten most significantly enriched pathway using DEGs, it was found that cancer-related signaling pathway such as VEGF and HIF-1 signaling pathways were up-regulated (Figure 2-19). Above all, *PXN* ($Log_2FC = 1.13$), *RAC2* ($Log_2FC = 1.34$), and *NOS3* ($Log_2FC = 1.12$) which promotes cell migration incorporating focal adhesion turnover and actin reorganization were up-regulated through VEGF signaling pathway (Figure 2-20). Interestingly, the upstream receptor of this pathway, *KDR* ($Log_2FC = -1.37$)

that encodes VEGFR2 which is one of the anti-cancer therapeutic target was decreased as previous report [100, 197]. Furthermore, thyroid hormone synthesis pathway as well as several metabolism pathways were down-regulated as *RAS*-positive ATC.

DNA methylation pattern of anaplastic thyroid carcinoma

To investigate genome-wide DNA methylation change during progression of ATC from WDTC, the analysis were conducted incorporating Illunima's HumanMethylation450 BeadChip data of TCGA [39]. In total, 158 samples including 50 normal thyroid tissues were applied to the analysis in addition to 16 ATC samples. Prior to integrated analysis, the possible batch effect between TCGA's methylation array and *HpaII* digestion WGS data was determined. Since the global DNA methylation profile also displayed distinction between *BRAF*-positive PTCs and the others as transcriptome profile (Figure 2-21A), it was supposed that CpG sites which do not show difference in DNA methylation level across all TCGA samples including normal thyroid tissues might be also have similar DNA methylation level to those of ATC. There were 352 CpGs which have exactly same DNA methylation level across 158 TCGA samples. As expected, 352 CpG positions of 16 ATC samples displayed highly concordance DNA methylation level to those of TCGA samples (Spearman's rank correlation coefficient $R^2 = 0.78$ on average; Figure 2-21B). Therefore, it was concluded that there is not huge batch effect between TCGA's methylation array and *HpaII* digestion WGS

data. The subsequent analyses were conducted using 32,408 CpG positions which have all DNA methylation level across analyzed samples.

Initially, PCA was conducted to get a glance whether the distinction between ATC and WDTC samples in DNA methylation level is present. As PCA result of transcriptome profile, it suggested that the global DNA methylation pattern of ATC is distinguished from WDTC (Figure 2-22A). Also, there was still separation within PTC samples according to the type of driver gene alteration. Then, the most variable 1,000 CpG sites across ATC and PTC was visualized to clarify the detailed difference of them. K-means clustering (K=4) was performed and it was found ATC specific hyermethylation cluster (cluster 3; Figure 2-22B). Particularly, CpG sites in this cluster were enriched in CpG islands (Chi-square test $p < 0.001$). It is known that DNA methyltransferase genes maintain or induce CpG islands hypermethylation and are involved in cancer progression [198-200]. It could be observed that the maintenance DNA methyltransferase, *DNMT1* ($Log_2FC = 1.09$), and *de novo* DNA methyltransferase, *DNMT3B* ($Log_2FC = 1.45$), were overexpressed in ATC and these activations might be contributed to alteration of DNA methylation pattern (Figure 2-23).

Table 2-2. The whole-genome sequencing summary of 16 ATC tumors.

Sample ID	Mean depth (X)	Total read	Total mapped read	Paired in sequencing	Read1	Read2
ATC1-T	72.54	1,426,795,397	1,416,984,853	1,426,795,397	712,178,142	714,617,255
ATC2-T	73.13	1,434,727,061	1,426,895,181	1,434,727,061	716,242,461	718,484,600
ATC3-T	83.16	1,656,406,324	1,641,798,355	1,656,406,324	826,403,247	830,003,077
ATC4-T	78.65	1,542,983,925	1,533,268,758	1,542,983,925	770,122,598	772,861,327
ATC5-T	47.90	938,825,110	933,826,893	938,825,110	468,800,729	470,024,381
ATC6-T	73.44	1,449,542,648	1,434,230,882	1,449,542,648	723,515,971	726,026,677
ATC7-T	91.63	1,785,925,488	1,778,133,997	1,785,925,488	891,987,586	893,937,902
ATC8-T	64.65	1,285,736,462	1,271,800,254	1,285,736,462	641,735,650	644,000,812
ATC9-T	60.50	1,186,318,091	1,181,369,160	1,186,318,091	592,249,119	594,068,972
ATC10-T	62.67	1,227,710,383	1,221,955,080	1,227,710,383	612,981,276	614,729,107
ATC11-T	80.46	1,579,242,347	1,569,733,294	1,579,242,347	788,156,301	791,086,046
ATC12-T	76.31	1,497,005,506	1,487,015,765	1,497,005,506	747,468,504	749,537,002
ATC13-T	101.07	1,988,961,527	1,981,127,415	1,988,961,527	993,005,515	995,956,012
ATC14-T	90.29	1,771,189,369	1,757,804,252	1,771,189,369	884,127,739	887,061,630
ATC15-T	79.34	1,565,465,455	1,559,615,416	1,565,465,455	781,116,251	784,349,204
ATC16-T	60.27	1,179,710,492	1,175,747,705	1,179,710,492	589,043,405	590,667,087

Table 2-3. The whole-genome sequencing summary of 15 matched normal samples.

Sample ID	Mean depth (X)	Total reads	Total mapped reads	Paired in sequencing	Read1	Read2
ATC1-N	36.48	710,040,415	708,192,645	710,040,415	354,811,320	355,229,095
ATC2-N	32.28	629,159,221	626,908,417	629,159,221	314,341,936	314,817,285
ATC3-N	37.16	731,393,916	728,432,772	731,393,916	365,415,787	365,978,129
ATC4-N	34.95	680,775,148	678,606,279	680,775,148	340,150,258	340,624,890
ATC5-N	40.27	783,041,894	780,334,431	783,041,894	391,210,834	391,831,060
ATC6-N	45.15	879,361,539	876,136,443	879,361,539	439,321,298	440,040,241
ATC7-N	39.30	764,651,735	761,434,491	764,651,735	381,943,711	382,708,024
ATC8-N	32.59	640,484,371	638,382,496	640,484,371	320,031,729	320,452,642
ATC9-N	34.38	670,597,312	668,686,518	670,597,312	334,972,767	335,624,545
ATC10-N	34.25	666,780,846	664,683,063	666,780,846	333,134,706	333,646,140
ATC11-N	34.11	668,174,283	665,382,789	668,174,283	333,764,263	334,410,020
ATC13-N	46.06	903,151,443	899,603,493	903,151,443	451,113,345	452,038,098
ATC14-N	37.44	729,837,882	727,298,031	729,837,882	364,517,198	365,320,684
ATC15-N	40.16	792,508,132	789,653,642	792,508,132	395,691,002	396,817,130
ATC16-N	28.16	552,775,431	550,146,300	552,775,431	275,973,641	276,801,790

Table 2-4. The RNA sequencing summary of 16 ATC tumors.

Sample ID	Total reads	Uniquely mapped reads	Uniquely mapped reads (%)
ATC1-T	51,255,995	47,355,152	94.41
ATC2-T	72,218,929	67,270,000	94.23
ATC3-T	87,848,849	80,816,219	93.43
ATC4-T	65,852,340	61,130,623	93.96
ATC5-T	39,131,827	33,909,833	87.55
ATC6-T	86,419,675	80,184,434	94.44
ATC7-T	54,360,687	49,663,536	92.39
ATC8-T	62,849,512	58,628,342	94.40
ATC9-T	62,929,267	58,167,329	93.53
ATC10-T	56,257,590	52,549,778	94.41
ATC11-T	63,513,300	58,920,203	93.83
ATC12-T	72,380,435	67,365,679	94.05
ATC13-T	49,356,820	42,608,481	87.55
ATC14-T	70,354,278	65,450,564	94.20
ATC15-T	57,551,365	52,013,710	91.94
ATC16-T	70,267,621	60,695,643	91.00

Table 2-5. The MAF of germline *TP53* fmutations in patients with various types of thyroid carcinoma.

Chr	Position	Ref	Alt	Amino acid Change	rsID	COSMIC	MAF					
							ExAC ALL	ExAC EAS	ATC normal	wiFTC normal	FTC normal	PTC normal
chr17	7,579,882	C	G	E11Q	rs201382018	COSM11606	0.004% (4/118,912)	0.05% (4/8,534)	3.13% (1/32)	5.56% (1/18)	0% (0/36)	0% (0/1,098)
chr17	7,578,388	C	T	R49H	rs397514495	COSM1386800	0.002% (2/121,316)	0% (0/8,652)	3.13% (1/32)	0% (0/18)	0% (0/36)	0% (0/1,098)

Abbreviations: MAF, minor allele frequency; ExAC, exome aggregative consortium; EAS, East Asian.

Table 2-6. The detailed annotation of somatic mutations of each sample.

Sample ID	Chr	Start	End	Ref	Alt	Gene	Substitution type	Amino acid change	COSMIC ID
ATC1-T	chr7	140,453,136	140,453,136	A	T	*BRAF*	missense SNV	p.V600E	COSM476
ATC2-T	chr7	140,453,136	140,453,136	A	T	*BRAF*	missense SNV	p.V600E	COSM476
	chr21	44,524,456	44,524,456	G	A	*U2AF1*	missense SNV	p.S34F	COSM166866
	chr5	1,295,228	1,295,228	C	T	*TERT*	promoter SNV	C228T	
ATC3-T	chr10	89,720,805	89,720,808	CTTT	-	*PTEN*	frameshift deletion	p.T319fs	COSM4896
	chr2	178,098,803	178,098,803	C	G	*NFE2L2*	missense SNV	p.G65A	
	chr5	1,295,228	1,295,228	G	A	*TERT*	promoter SNV	C228T	
ATC4-T	chr7	140,453,136	140,453,136	A	T	*BRAF*	missense SNV	p.V600E	COSM476
	chrX	76,939,606	76,939,606	G	T	*ATRX*	stopgain SNV	p.S343X	
	chr5	1,295,228	1,295,228	G	A	*TERT*	promoter SNV	C228T	
ATC5-T	chr7	140,453,136	140,453,136	A	T	*BRAF*	missense SNV	p.V600E	COSM476
	chr17	7,578,395	7,578,395	G	A	*TP53*	missense SNV	p.H47Y	COSM129849
	chr5	1,295,228	1,295,228	G	A	*TERT*	promoter SNV	C228T	
	chr5	1,295,250	1,295,250	G	A	*TERT*	promoter SNV	C250T	
ATC6-T	chr10	70,405,936	70,405,936	A	-	*TET1*	frameshift deletion	p.T1150fs	
ATC7-T	chr11	533,874	533,874	T	C	*HRAS*	missense SNV	p.Q61R	COSM499
	chr17	7,577,565	7,577,565	T	-	*TP53*	frameshift deletion	p.N107fs	
	chrX	20,148,726	20,148,726	C	A	*EIF1AX*	splicing site		

Table 2-6. continued.

Sample ID	Chr	Start	End	Ref	Alt	Gene	Substitution type	Amino acid change	COSMIC ID
ATC7-T	chr5	1295228	1295228	G	A	*TERT*	promoter SNV	C228T	
ATC8-T	chr11	534,286	534,286	C	G	*HRAS*	missense SNV	p.G13R	COSM99938
	chr14	105,246,551	105,246,551	C	T	*AKT1*	missense SNV	p.E17K	COSM33765
	chr17	7,574,000	7,574,000	C	A	*TP53*	stopgain SNV	p.E211X	COSM11078
	chrX	20,148,727	20,148,727	T	C	*EIF1AX*	splicing site		
ATC9-T	chr7	140,453,136	140,453,136	A	T	*BRAF*	missense SNV	p.V600E	COSM476
	chr6	150,001,515	150,001,515	T	-	*LATS1*	frameshift deletion	p.R697fs	
	chr13	21,549,393	21,549,393	C	-	*LATS2*	frameshift deletion	p.R961fs	
	Chr5	1,295,228	1,295,228	G	A	*TERT*	promoter SNV	C228T	
ATC10-T	chr17	7,577,120	7,577,120	C	T	*TP53*	missense SNV	p.R141H	COSM99729
ATC11-T	chr1	115,256,529	115,256,529	T	C	*NRAS*	missense SNV	p.Q61R	COSM584
	chr17	41276133	41276133	C	G	*BRCA1*	splicing site		
	chrX	20,156,734	20,156,734	C	T	*EIF1AX*	missense SNV	p.G8E	
	chr5	1,295,250	1,295,250	G	A	*TERT*	promoter SNV	C250T	
ATC12-T	chr7	140,453,136	140,453,136	A	T	*BRAF*	missense SNV	p.V600E	COSM476
	chr14	105,246,551	105,246,551	C	T	*AKT1*	missense SNV	p.E17K	COSM33765
	chr5	1,295,228	1,295,228	G	A	*TERT*	promoter SNV	C228T	
ATC13-T	chr1	115,256,528	115,256,528	T	C	*NRAS*	missense SNV	p.Q61R	COSM584

Table 2-6. continued.

Sample ID	Chr	Start	End	Ref	Alt	Gene	Substitution type	Amino acid change	COSMIC ID
ATC13-T	chrX	20,156,734	20,156,734	C	T	*EIF1AX*	missense SNV	p.G8E	
	chr5	1,295,228	1,295,228	G	A	*TERT*	promoter SNV	C228T	
ATC14-T	chr1	115,256,529	115,256,529	T	C	*NRAS*	missense SNV	p.Q61R	COSM584
	chrX	20,156,734	20,156,734	C	T	*EIF1AX*	missense SNV	p.G8E	
	chr5	1,295,228	1,295,228	G	A	*TERT*	promoter SNV	C228T	
ATC15-T	chr7	151,927,023	151,927,023	G	C	*KMT2C*	stopgain SNV	p.Y987X	COSM216053
	chr7	151,927,067	151,927,067	T	C	*KMT2C*	missense SNV	p.R973G	COSM4161994
ATC16-T	chr7	140,453,136	140,453,136	A	T	*BRAF*	missense SNV	p.V600E	COSM476
	chr3	178,952,085	178,952,085	A	G	*PIK3CA*	missense SNV	p.H1047R	COSM94986
	chr17	7,577,539	7,577,539	G	A	*TP53*	missense SNV	p.R116W	COSM3388183
	chr5	1,295,228	1,295,228	G	A	*TERT*	promoter SNV	C228T	

Abbreviations: Chr, chromosome; Ref, reference allele; Alt, alternative allele.

Table 2-7. The significantly altered chromosome arm in ATC.

Arm	# genes	Amp frequency	Amp *q*-value	Del frequency	Del *q*-value
1p	2121	0.08	0.932	0.08	0.86
1q	1955	0.23	0.424	0.09	0.86
2p	924	0	0.932	0.07	0.86
2q	1556	0	0.932	0.07	0.86
3p	1062	0.17	0.885	0.17	0.86
3q	1139	0.23	0.523	0.09	0.86
4p	489	0	0.932	0.36	0.227
5p	270	0.31	0.424	0.1	0.86
5q	1427	0.23	0.481	0.09	0.86
6p	1173	0.08	0.932	0.08	0.86
6q	839	0	0.932	0.29	0.472
7p	641	0.25	0.523	0.18	0.86
7q	1277	0.17	0.885	0.17	0.86
8p	580	0.18	0.885	0.25	0.818
8q	859	0.23	0.525	0.09	0.86
9p	422	0.09	0.932	0.23	0.86
9q	1113	0.09	0.932	0.23	0.818
10p	409	0	0.932	0.14	0.86
10q	1268	0	0.932	0.14	0.86
11p	862	0.14	0.932	0	0.919
11q	1515	0.14	0.885	0	0.919
12p	575	0.08	0.932	0.08	0.86
12q	1447	0.08	0.932	0.15	0.86
13q	654	0	0.932	0.14	0.86
14q	1341	0.08	0.932	0.08	0.86
15q	1355	0	0.932	0.21	0.818
16p	872	0.08	0.932	0.15	0.86
16q	702	0.15	0.932	0.08	0.86
17p	683	0	0.932	0.43	0.0567
17q	1592	0.23	0.46	0.09	0.86
18p	143	0.09	0.932	0.23	0.86
18q	446	0.08	0.932	0.15	0.86
19p	995	0.23	0.523	0.09	0.86

Table 2-7. continued.

Arm	# genes	Amp frequency	Amp *q*-value	Del frequency	Del *q*-value
19q	1709	0.29	0.178	0	0.919
20p	355	0.31	0.424	0.1	0.86
20q	753	0.36	0.153	0	0.919
21p	13	0.77	1.89E-07	0.25	0.86
21q	509	0.5	0.0225	0.44	0.148
22q	921	0.1	0.932	0.31	0.377

Abbreviations: Amp, amplification; Del, deletion.

Table 2-8. The regions with significant amplification in ATC.

Cytoband	Wide peak boundaries	Genes in wide peak	Residual *q*-value[a]
2q21.2	chr2:133,034,593-133,118,006	*ANKRD30BL*	0.0001
16q22.2	chr16:70,882,524-71,200,029	*HYDIN*	0.0001
6p11.2	chr6:57,212,739-57,568,050	*hsa-mir-548u, PRIM2*	0.0001
15q26.3	chr15:102,504,813-102,531,392	*WASH3P, DDX11L1, DDX11L9*	0.0001
1p36.13	chr1:16,837,477-16,838,671	*CROCCP3*	0.0007
1q21.1	chr1:145,374,344-145,380,272	*NBPF10*	0.0011
7q11.22	chr7:71,563,279-71,589,549	*CALN1*	0.0047
17q12	chr17:36,269,518-36,405,553	*TBC1D3F, LOC440434, TBC1D3*	0.0047
2p11.1	chr2:91,676,990-91,823,234	*LOC654342*	0.0051
14q11.2	chr14:19,385,975-20,162,672	*POTEG, POTEM, LOC642426*	0.0093
10q11.22	chr10:47,021,072-47,104,171	*PPYR1, LOC643650*	0.0216

[a] The residual *q*-value denotes the significance of peak region after excluding overlapped amplifications and deletions in other, more significant peak regions.

Table 2-9. The regions with significant deletion in ATC.

Cytoband	Wide peak boundaries	Genes in wide peak	Residual *q*-value[a]
9p21.3	chr9:22,002,865-22,009,400	*CDKN2A, CDKN2B*	5.88E-07
2q36.1	chr2:223,808,117-224,461,675	*KCNE4*	0.0004
19p13.3	chr19:1,279,242-1,355,035	*ENFA2*	0.0194
8p23.1	chr8:7,439,997-7,946,469	*DEFB4A, SPAG11B, DEFB103B, DEFB104A, DEFB105A, DEFB106A, DEFB107A, DEFB103A, FAM90A13, FAM90A8, FAM90A18, FAM90A9, FAM90A10, DEFB107B, DEFB104B, DEFB106B, DEFB105B, DEFB109P1B, FAM90A14, SPAG11A, FAM90A19, FAM66E, LOC100132396*	0.0194
22q13.32	chr22:48,934,669-48,943,223	*LOC284933*	0.0194
9q22.2	chr9:92,221,450-92,782,978	*UNQ6494*	0.0283
17p11.2	chr17:21,826,480-22,022,469	*FLJ36000*	0.0213
21q22.3	chr21:46,046,559-46,048,316	*KRTAP10-9*	0.0371

[a] The residual *q*-value denotes the significance of peak region after excluding overlapped amplifications and deletions in other, more significant peak regions.

Table 2-10. The copy number status of *CDNK2A*/*CDKN2B* in ATC cell lines from CCLE.

	Cell line	CAL62	SW579	BHT101	8305C	8505C	BCPAP	FTC133	FTC238	ML1	TT2609C02	TT
	Histology	ATC	ATC	ATC	ATC	ATC	PTC	FTC	FTC	FTC	FTC	MTC
Copy number values	*CDKN2A*	-4.67[a]	-2.84	-0.55	0.42	0.49	-0.58	-0.12	0.31	-0.63	-0.72	-0.96
	CDKN2B	-4.74	-4.94	-0.55	0.42	0.49	-0.58	-0.12	0.31	-0.63	-0.72	-0.96
Gene expression level	*CDKN2A*	0.02[b]	0.02	23.00	20.50	10.42	9.64	43.45	33.70	11.27	15.87	0.18
	CDKN2B	0.00	0.01	35.13	3.56	2.21	2.71	6.93	14.52	9.79	2.25	0.25

[a] Each value denotes $\log_2$ ratios (specifically, $\log_2$(copy number/2)) from circular binary segmentation algorithm.
[b] Each value denotes RPKM.

Table 2-11. The gene expression level of potential immunotherapeutic targets of cell lines from CCLE.

Cell line	CAL62	SW579	BHT101	8305C	8505C	BCPAP	FTC133	FTC238	ML1	TT2609C02	TT
Histology	ATC	ATC	ATC	ATC	ATC	PTC	FTC	FTC	FTC	FTC	MTC
IL6	0.69[a]	0.62	5.62	12.31	4.66	53.24	3.60	0.50	3.59	3.47	0.45
CD274 (PD-L1)	32.49	0.58	1.48	17.80	48.14	13.84	1.80	20.22	1.17	0.71	1.03
PDCD1LG2 (PD-L2)	3.78	2.85	0.63	12.85	14.17	4.54	1.84	3.99	0.56	0.01	0.00

[a] Each value denotes RPKM.

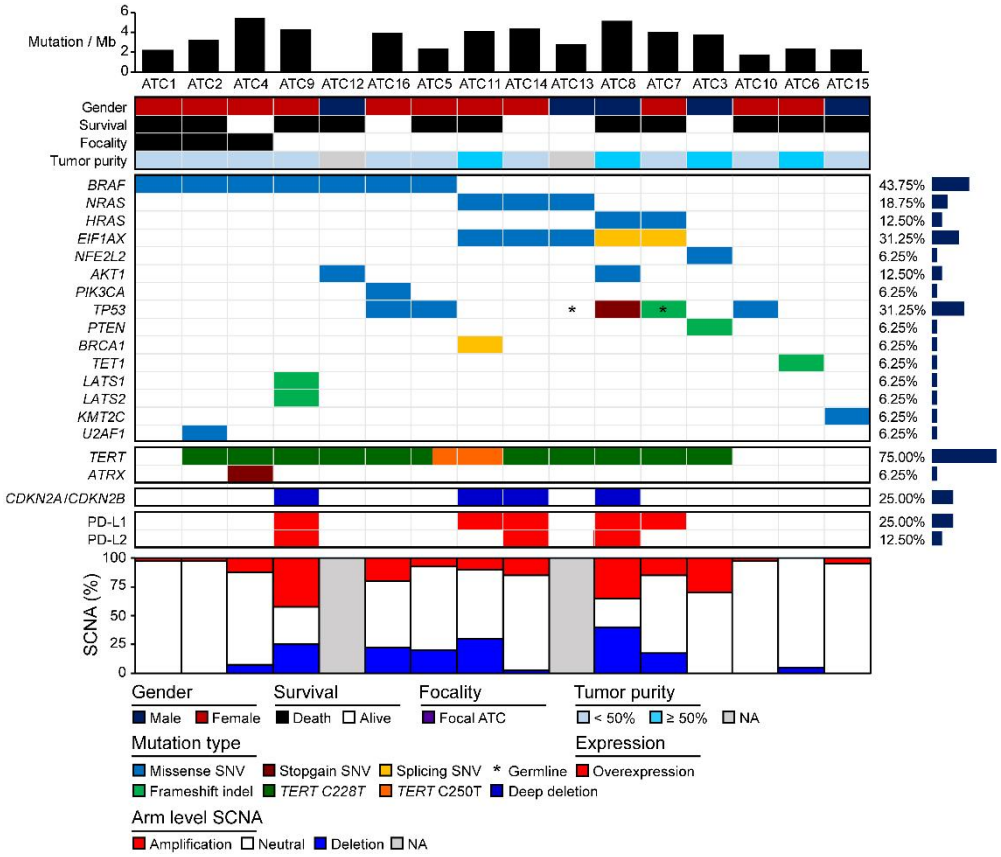

Figure 2-2. The genomic and transcriptomic alterations of 16 ATC samples. Each column represents individual sample.

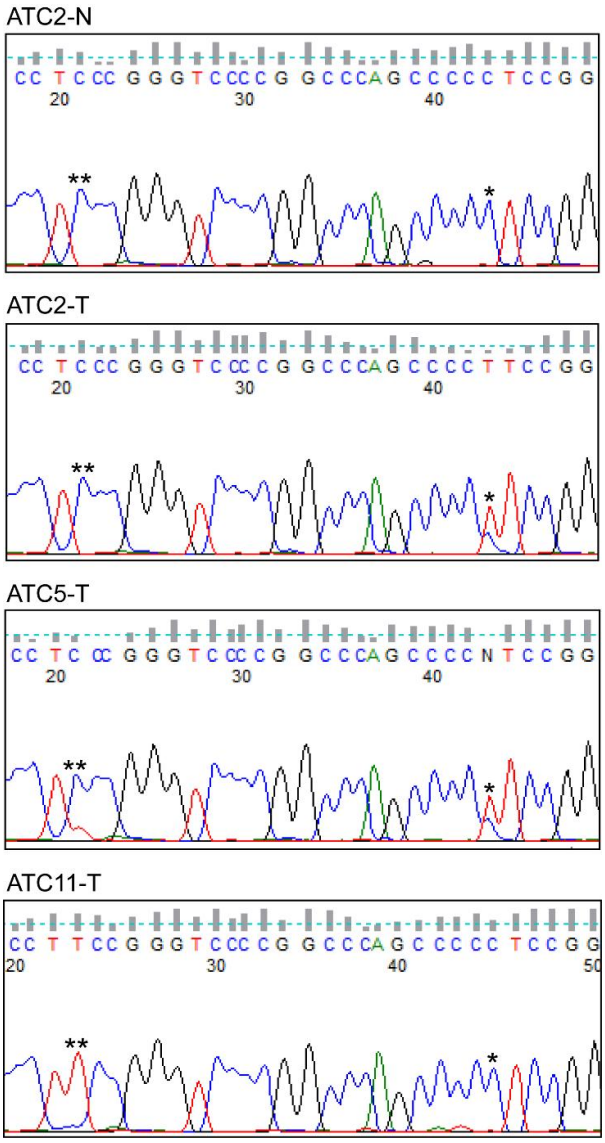

Figure 2-3. The representative results of *TERT* promoter sequencing. * and ** represent C228T and C250T, respectively.

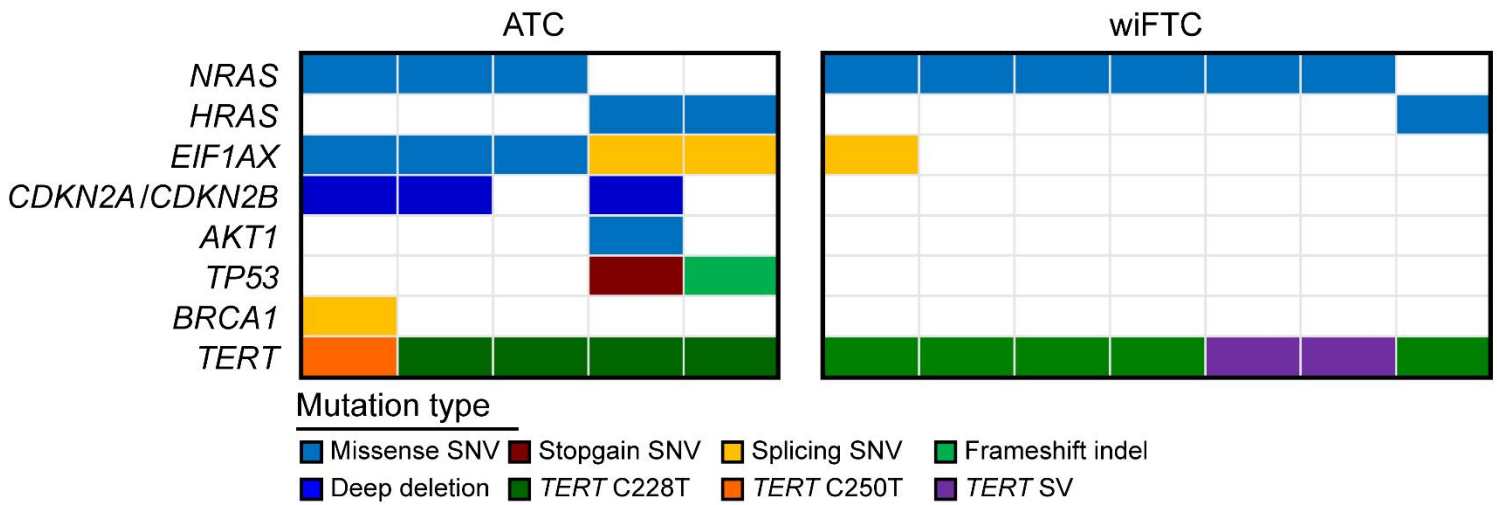

Figure 2-4. The comparison of mutation profiles between ATC and wiFTC. Each column represents individual sample.

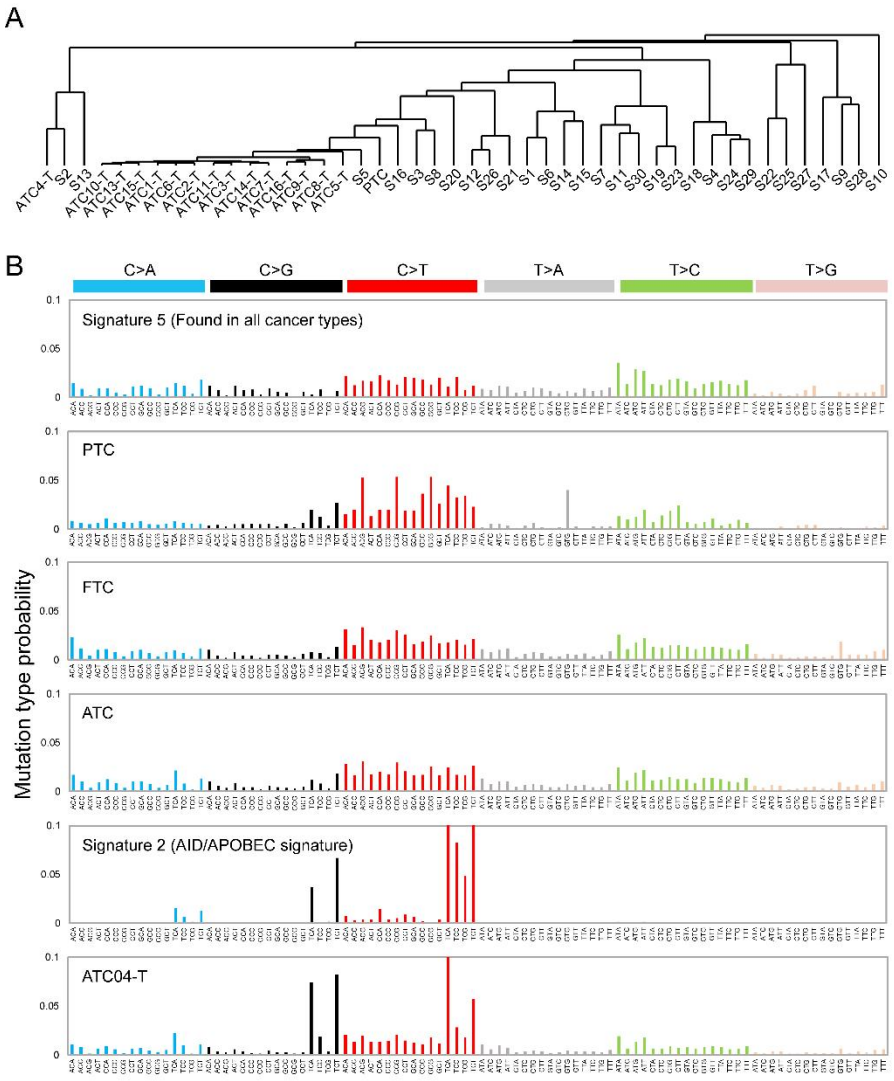

Figure 2-5. Mutational signature analysis. A) The mutational signature assignment of 15 ATC samples using average linkage hierarchical clustering. PTC denotes the average mutational signature across TCGA's whole-exome sequenced samples. S denotes reference signature from COSMIC. B) The representative mutational signatures of thyroid cancers. FTC represents the mutational signature of one whole-genome sequenced wiFTC sample. ATC represents the average mutational signature across 14 ATC samples except ATC4-T.

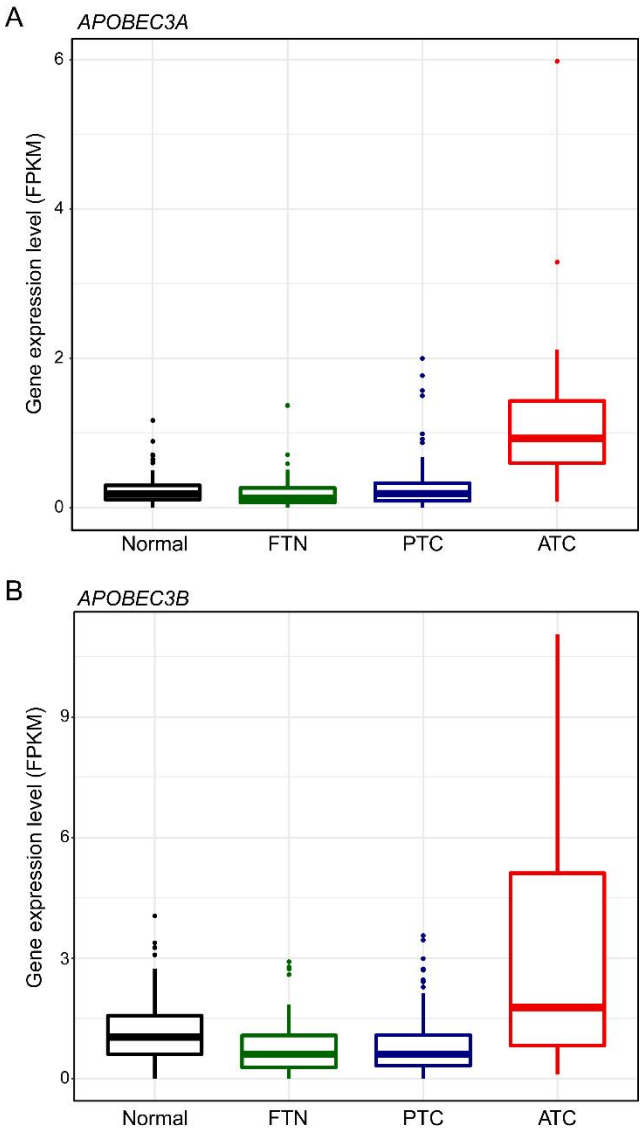

Figure 2-6. The overexpression of *APOBEC3A* and *APOBEC3B* in ATC. The expression level of A) *APOBEC3A* and B) *APOBEC3B* in normal thyroid tissue and various types of thyroid cancer.

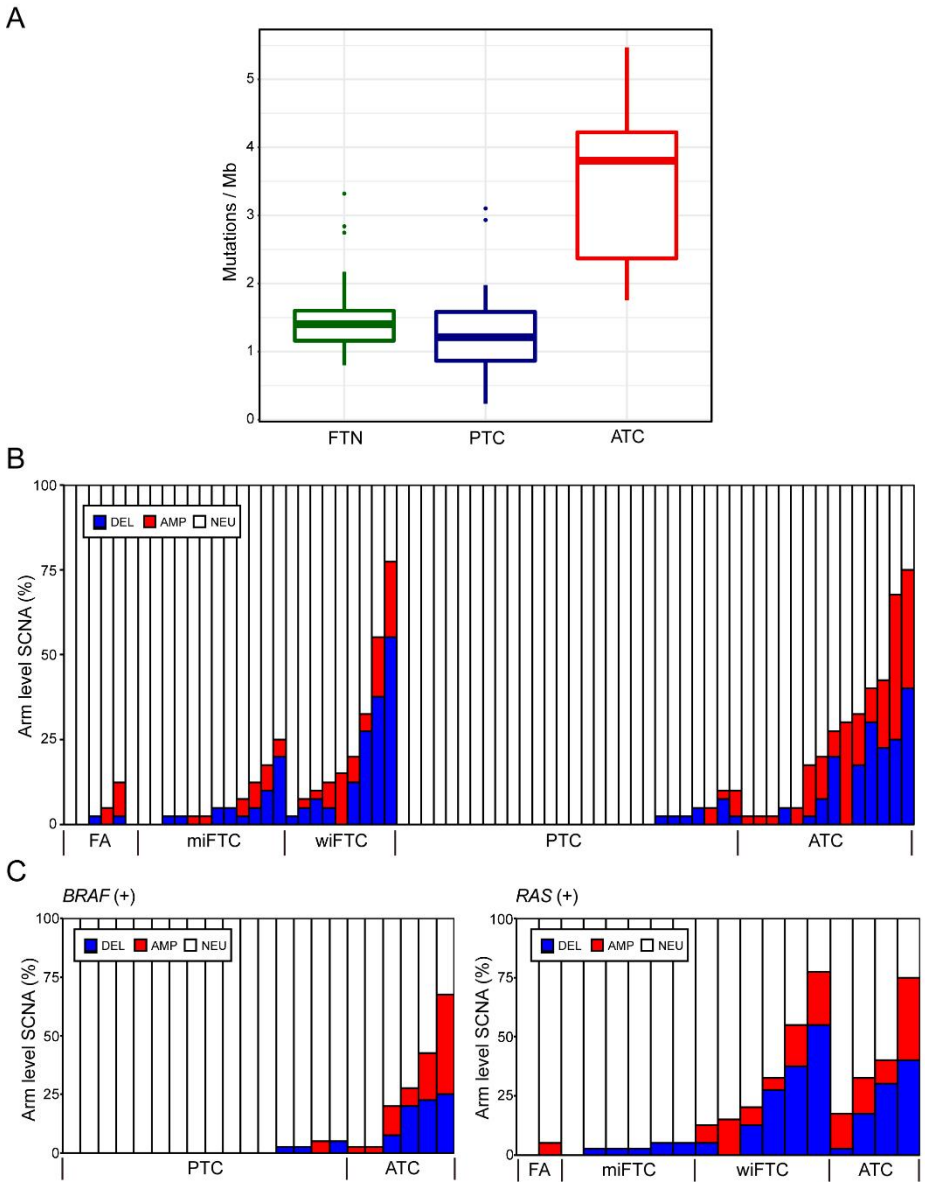

Figure 2-7. The aggressive genomic characteristics of ATC. A) The increment of mutation burden in ATC. B) The increment of arm-level SCNA in ATC. C) The comparison of arm-level SCNA in *BRAF*-positive and *RAS*-positive thyroid cancers, seperatively.

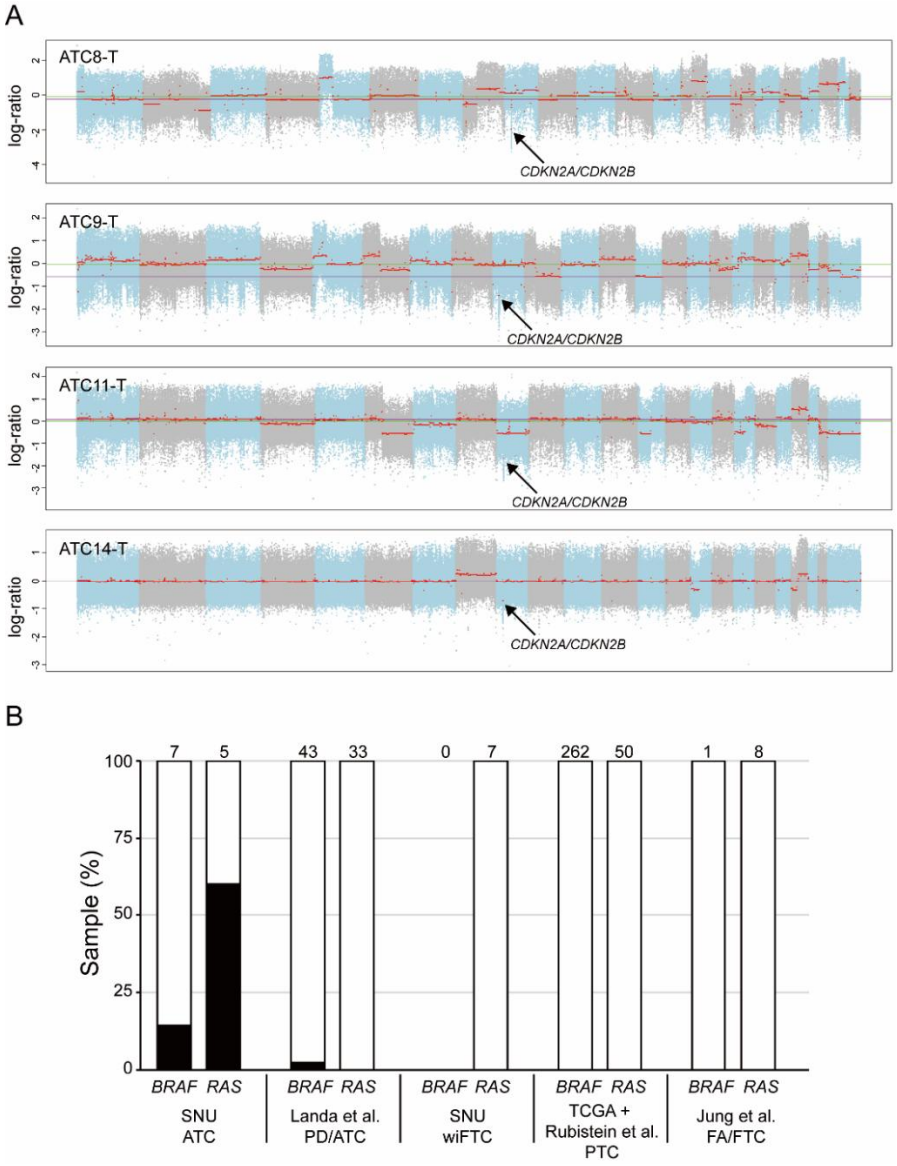

Figure 2-8. The recurrent *CDKN2A*/*CDKN2B* deletion in ATC. A) The genome-wide SCNA profile of four ATC samples. The region with *CDKN2A*/*CDKN2B* was indicated by arrows. B) The proportion of *CDKN2A*/*CDKN2B* deletion in several thyroid cancer studies. The shaded areas indicate the proportion of sample with *CDKN2A*/*CDKN2B* deletion. The numbers on the bar chart denote total sample with *BRAF* or *RAS* mutation.

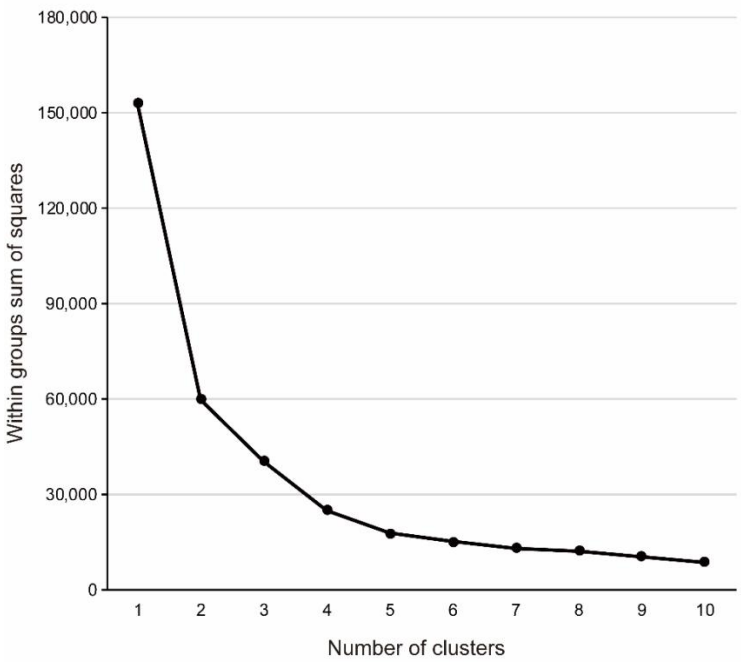

Figure 2-9. The determination of optimal number of cluster for molecular subtype classification using elbow method.

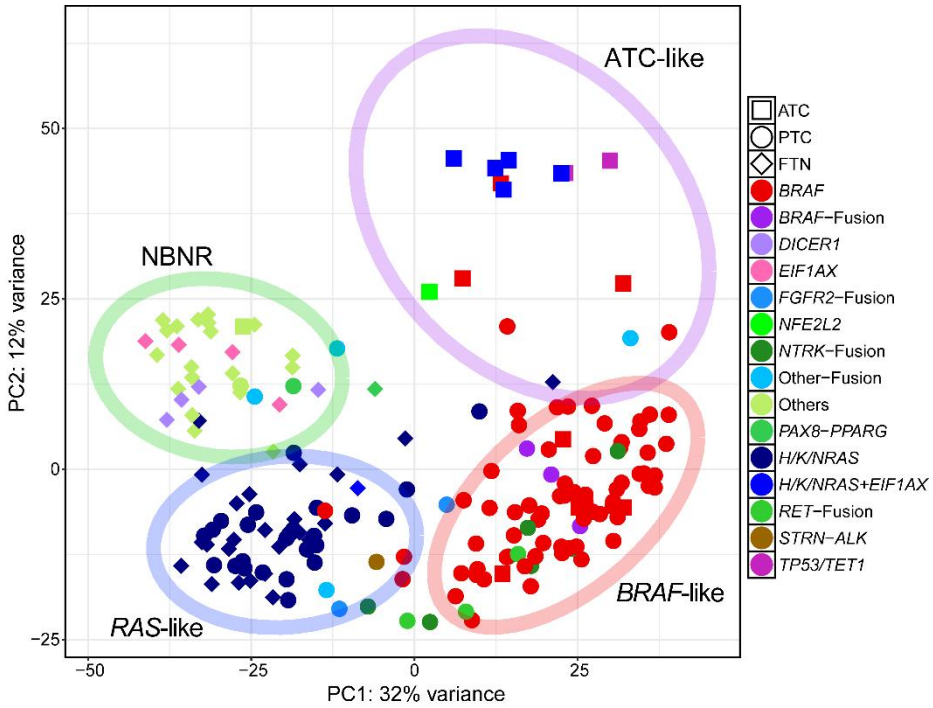

Figure 2-10. The molecular subtype classification of 16 ATC samples using K-means clustering via PCA (K=4). The most variable 500 genes within 16 ATC and 170 WDTC samples were used. The histological subtype and the driver mutation of each sample were represented by the shape and the color, respectively. 95.00% confidence ellipse displayed each cluster.

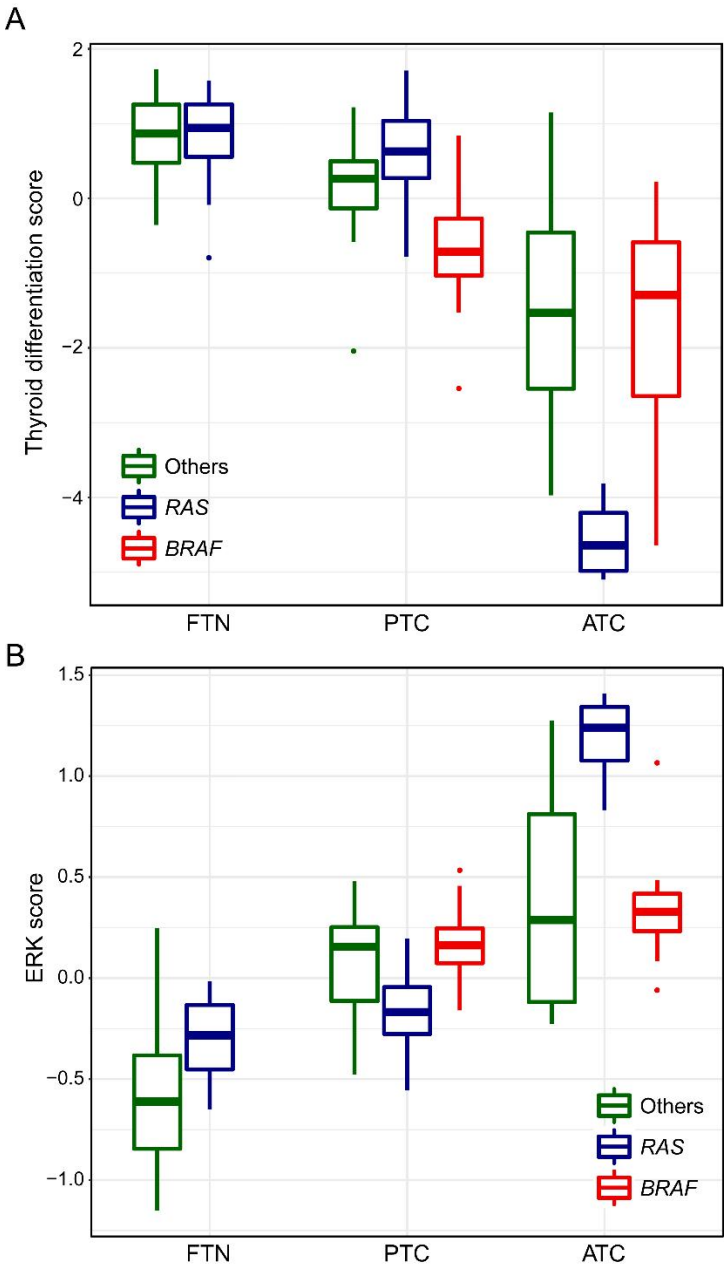

Figure 2-11. The analyses of thyroid differentiation loss and ERK/MAPK signaling pathway activation. using A) TDS and B) ERK score.

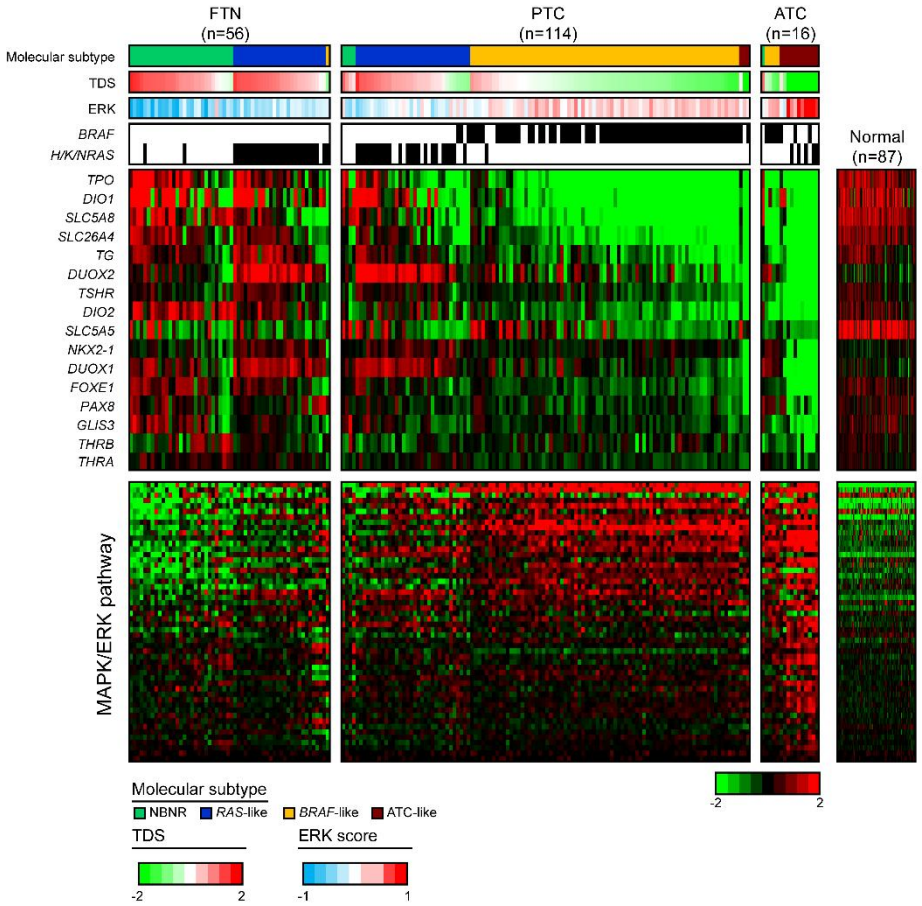

Figure 2-12. The expression profile of thyroid function and metabolism and MAPK/ERK pathway genes. 186 samples were sorted by the molecular subtype, and high to low TDS. Genes were sorted by low to high variability. Right heatmap displays expression level of normal thyroid tissues.

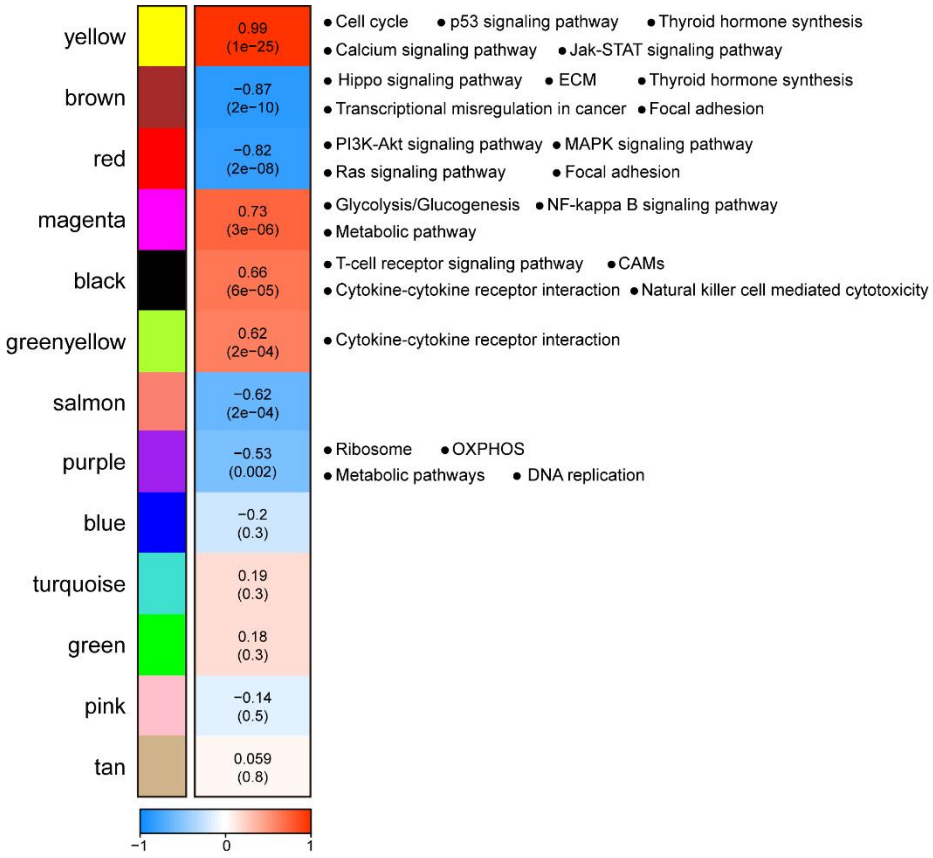

Figure 2-13. Weighted gene co-expression analysis using *RAS*-positive FTN and FTC origin ATC samples. Values outside and inside bracket indicate Pearson's correlation coefficient and *p*-value.

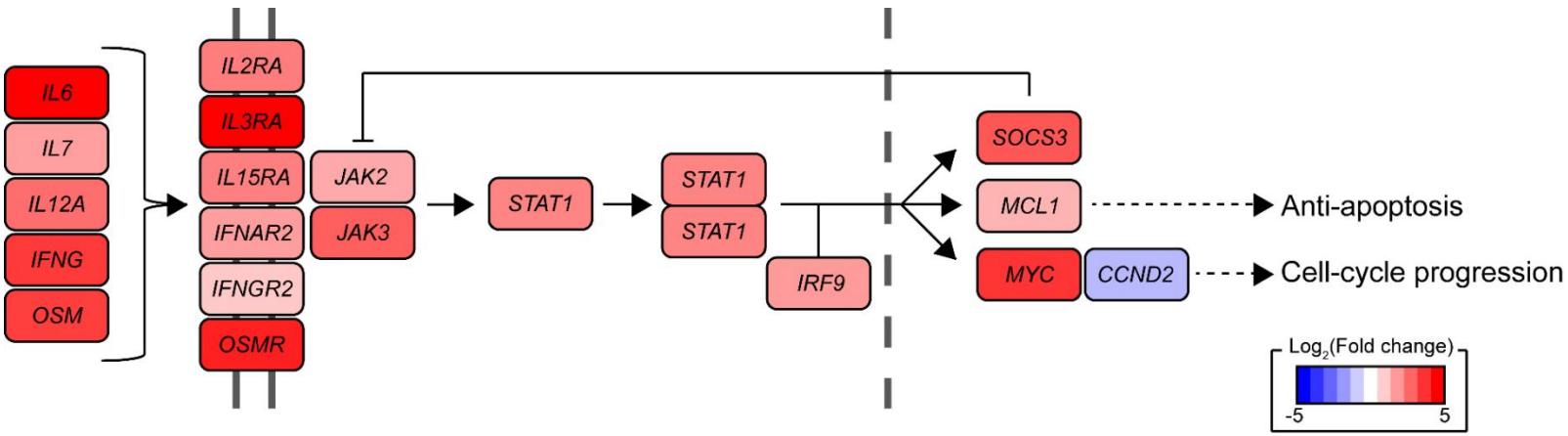

Figure 2-14. The activation of JAK/STAT pathway during progression of *RAS*-positive ATC from FTN. Each gene was colored by Log_2FC values of ATC relative to FTN.

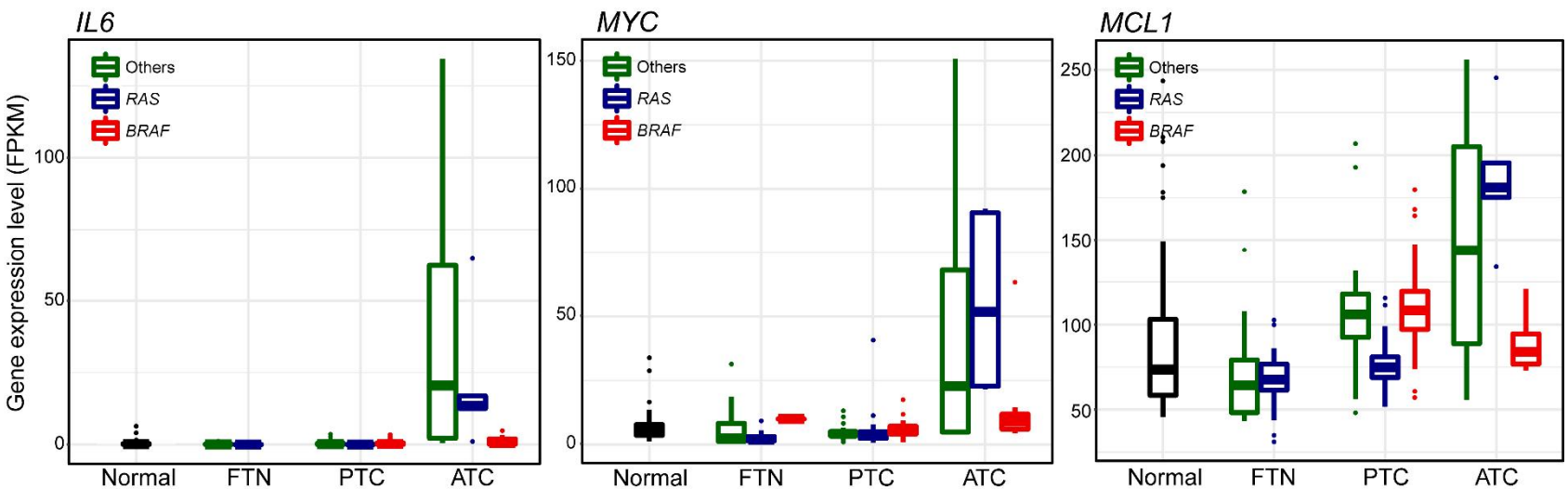

Figure 2-15. The IL6-dependent activation of *MYC* and *MCL1* in ATC.

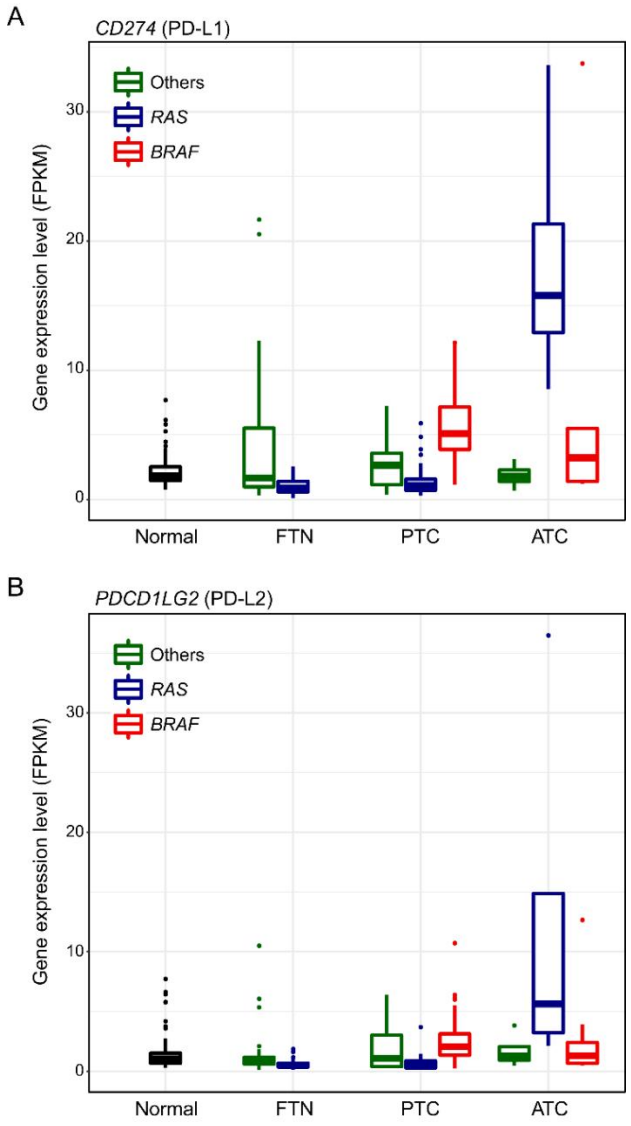

Figure 2-16. The activation of two immunotherapeutic targets in ATC. A) *CD274* (PD-L1) and B) *PDCD1LG2* (PD-L2).

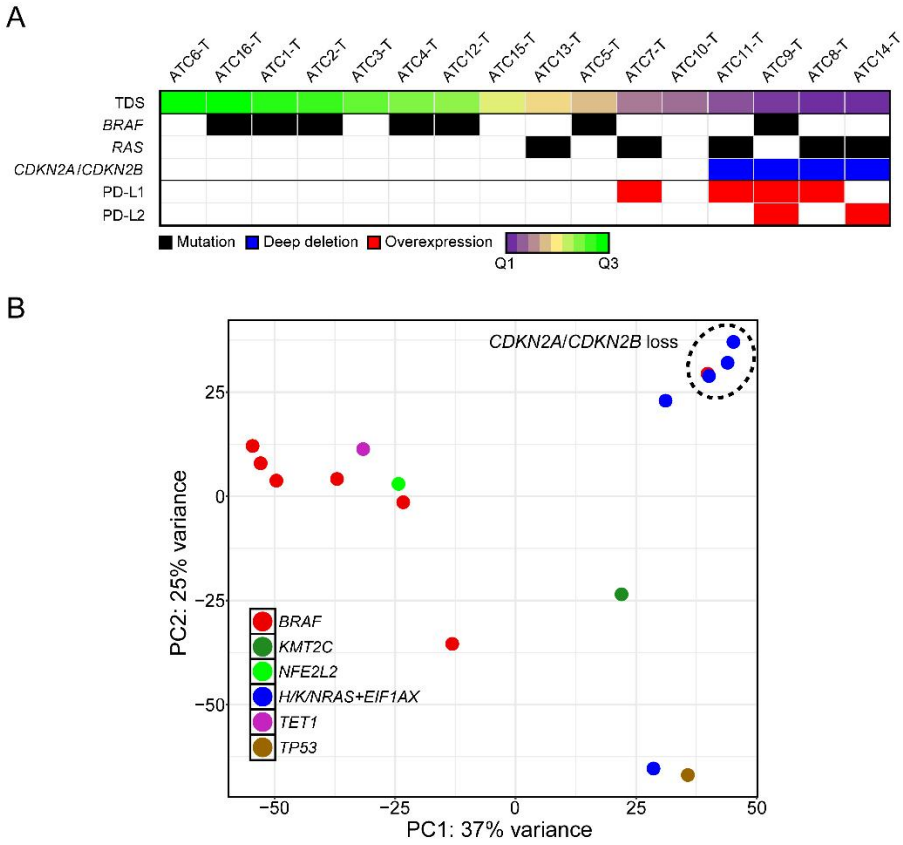

Figure 2-17. The relationship between *CDKN2A*/*CDKN2B* deletion and aggressiveness of ATC. A) The 16 ATC samples were sorted as high to low TDS. B) PCA using 16 ATC samples. Driver mutation of each sample were represented by the color

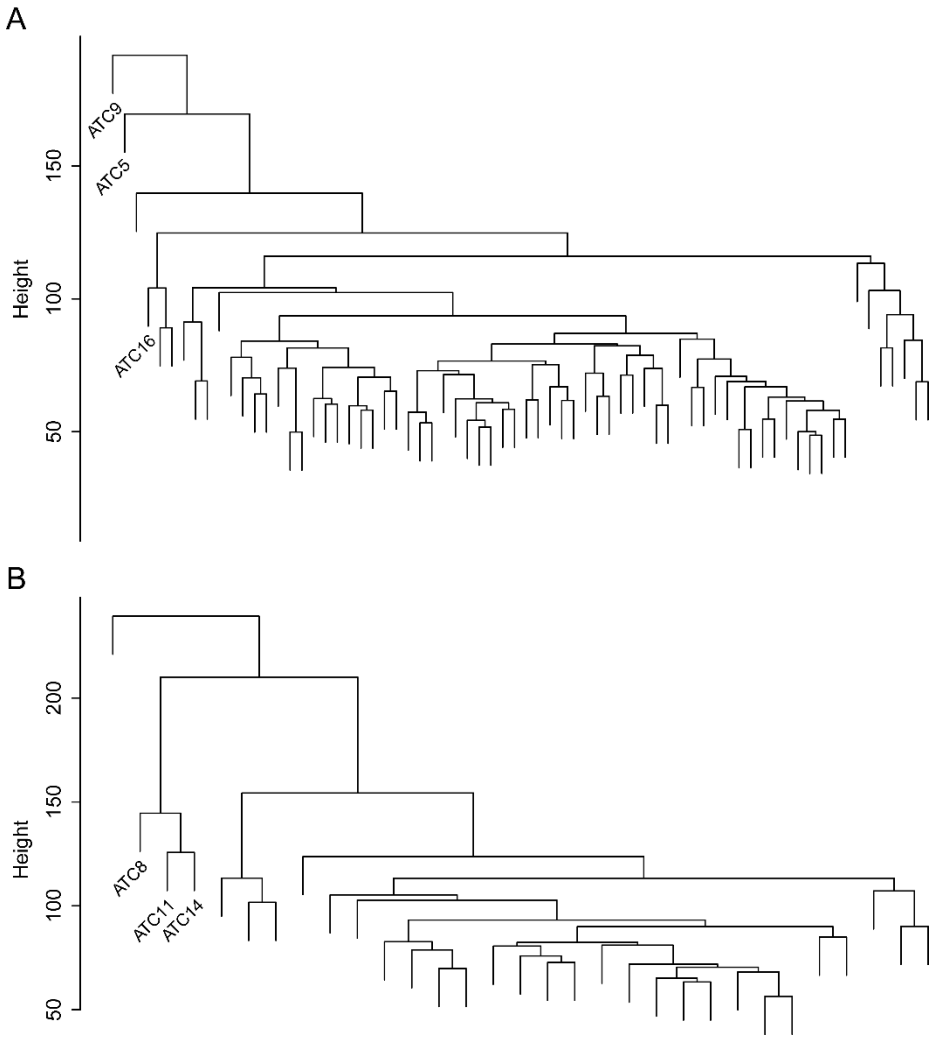

Figure 2-18. Determination of gene expression heterogeneity among *BRAF*-positive ATC. A) Analysis using *BRAF*-positive ATC and PTC samples. The unannotated braches represent PTC. B) Analysis using *RAS*-positive ATC and FTN. The unannotated braches represent FTN. The analyses were performed by average linkage hierarchical clustering

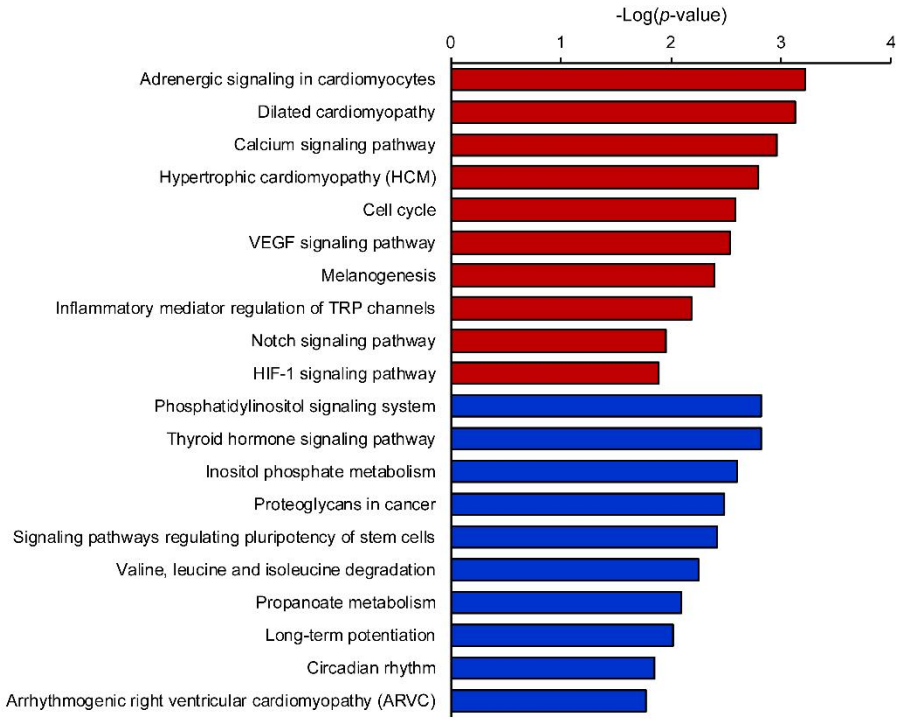

Figure 2-19. The result of pathway enrichment using up- and down-regulated genes in *BRAF*-positive ATC relative to *BRAF*-positive PTC. Red and blue bars indicate up- and down-regulated pathways, respectively.

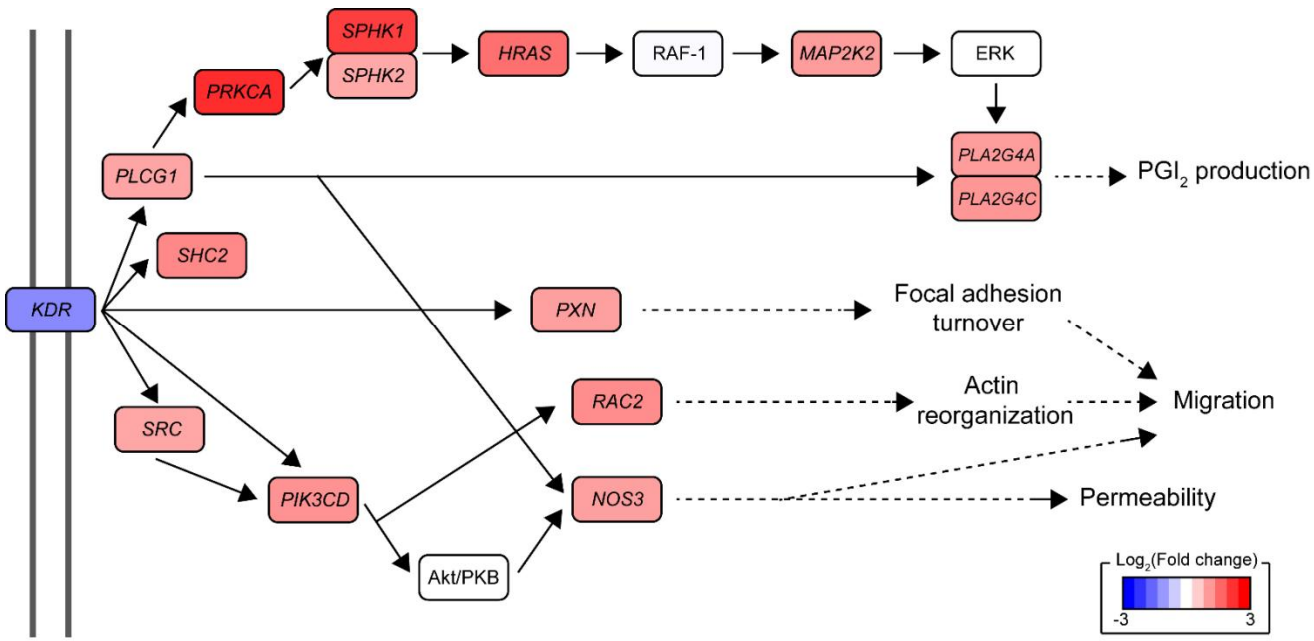

Figure 2-20. The activation of VEGF signaling pathway in *BRAF*-positive ATC. Each gene was colored by Log_2FC values of ATC relative to PTC.

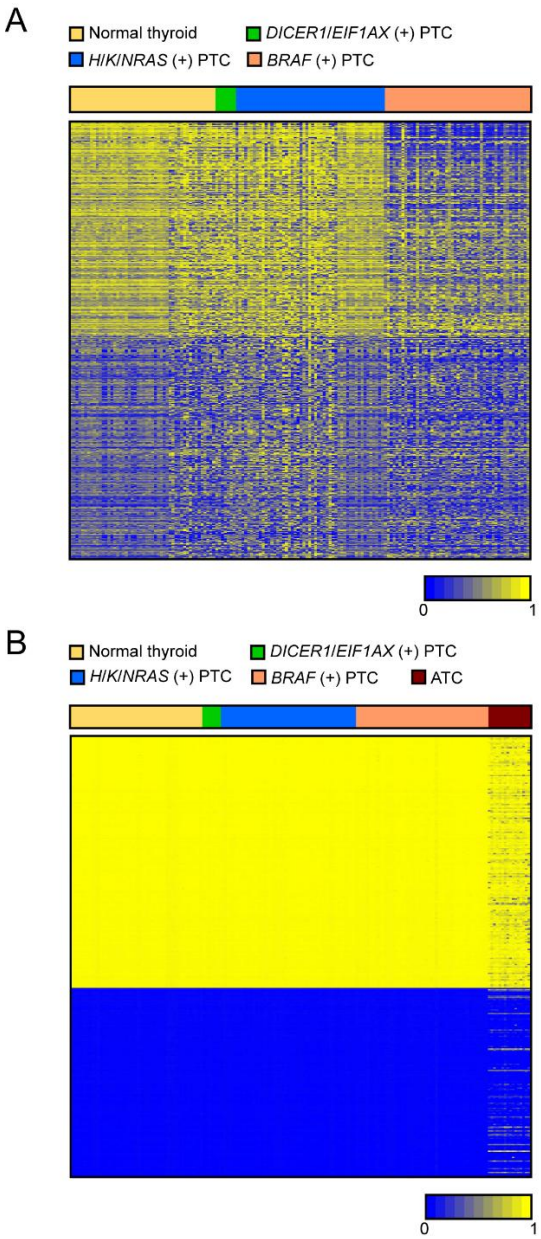

Figure 2-21. The determination of potential batch effect between TCGA's microarray and *HpaII* digestion WGS. A) The most variable 500 CpG positions across TCGA samples. B) 352 CpG positions which have exactly same DNA methylation level across TCGA samples were exhibited with 16 ATC samples.

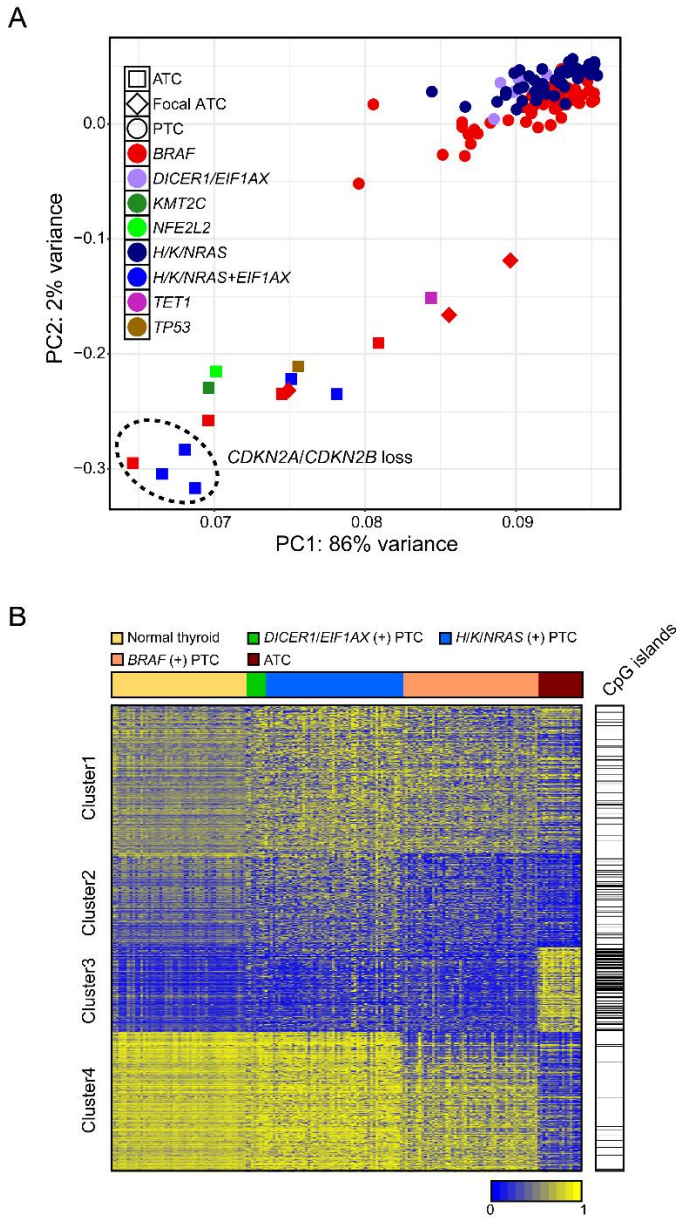

Figure 2-22. DNA methylation profile of ATC. A) PCA using 108 TCGA and 16 ATC samples. The histological subtype and the driver mutation of each sample were represented by the shape and the color, respectively. B) The most variable 1,000 CpG positions across TCGA and ATC samples were clustered by K-means clustering algorithm (K=4).

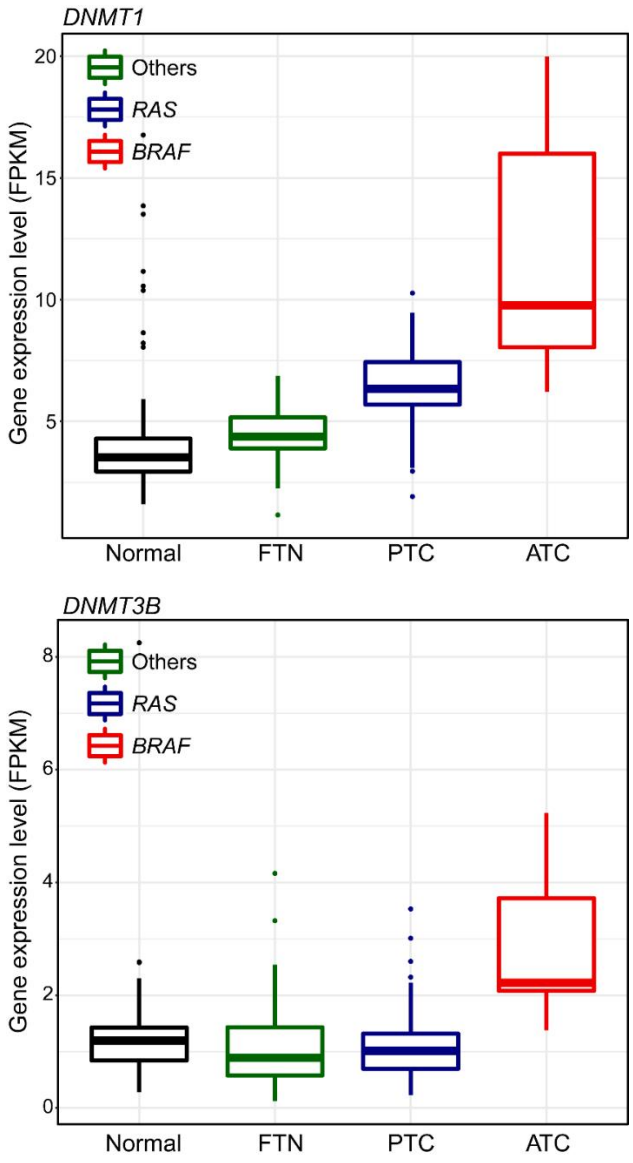

Figure 2-23. The activation of two methyltransferase genes in ATC.

Discussion

In recent years, with a rapid development of MPS technology, the molecular features of two major types of thyroid cancer, PTC and FTC, were determined by TCGA and the firsts chapter of this thesis [39]. These researches well illustrated the distinct mutational profile between PTC and FTC. Also, the novel classification of thyroid cancer based on the molecular level, rather than the traditional histological approach, was evaluated. However, the molecular feature of ATC which is the most lethal type of thyroid cancer is not still fully unveiled. Therefore, in this thesis, WGS and RNA-seq was performed on 16 ATC samples to understand the molecular pathogenesis of ATC.

Based on previous studies, it is now well-known that *BRAF* mutation and diverse types of fusion genes induce papillary histology and other driver genes such as *H/K/NRAS*, *DICER1*, *EIF1AX*, *IDH1*, and *PAX8-PPARG* contribute to follicular histology of thyroid cancer [39]. In 16 ATC samples, somatic mutations in *BRAF* and *H/NRAS* were discovered with the incidence of 43.75% and 31.25%, respectively (Figure 2-2). This result suggests that ATC could be advanced from PTC and FTC.

It is well established that *TERT* promoter mutations and SVs adjacent to *TERT* enhance the aggressiveness or poor prognosis of various types of cancer including WDTC [201-206]. Recently, the incidence of *TERT* promoter

mutation in ATC was reported as 73% [101]. It is in parallel with our result; *TERT* promoter mutations were identified in 75% of study subjects. Hence, our result again confirms the association between *TERT* and aggressiveness of thyroid cancer.

WDTC represents very mutually exclusive mutational profile; the majority of them have alteration of only one driver gene and do not accompany other mutations in other oncogene or TSG [39]. However, 75% of ATC samples had more than one mutation in cancer related genes in the present study (Figure 2-2). Interestingly, all five *RAS*-positive samples had co-mutation in *EIF1AX* which is recently revealed as a driver gene of thyroid cancer [39, 124]. It was reported that *RAS-EIF1AX* co-mutation is recurrently identified in PDTC and ATC [101]. Furthermore, three out of five *RAS-EIF1AX* co-mutated samples also harbored *CDKN2A/CDKN2B* loss in this study. This consequence was not reported in any other researches but this deletion was mostly confirmed in ATC among thyroid cancer cell lines (Table 2-10). Since ATC samples with *CDKN2A*/*CDKN2B* display the lowest thyroid cell differentiation, this alteration may play a crucial role in aggressiveness of ATC (Figure 2-17).

Moreover, somatic mutations in *NFE2L2*, *AKT1*, and *PIK3CA* were found in this study. *NFE2L2* is frequently mutated in lung squamous cell carcinoma [161], but the somatic mutation in this gene has never been reported in thyroid cancer. However, recent study discovered *PAX8-NFE2L2* fusion gene in PTC

[31]. Thus, *NFE2L2* could be suggested as a driver gene candidate in thyroid cancer.

TP53, *PTEN*, *TET1*, *LATS1*, and *LATS2* mutations in five ATCs were identified and most of them was truncating mutations (frameshift indel or stopgain SNV). *TP53* and *PTEN* are the most well-known TSGs and the truncation of *TET1*, *LATS1*, and *LATS2* are reported in several cancers but not in thyroid cancer yet [164, 165, 176, 177]. Furthermore, *U2AF1* and *KMT2C* which are the spliceosomal component and the histone lysine methyltransferase and these genes are known to be mutated in various types of cancer [166, 178, 179].

Intriguingly, germline mutations were found in *TP53* in two ATC patients with more frequent incidence than patients with WDTC. It is reported that germline *TP53* mutation contribute to early onset cancer development [207, 208], but two patients with germline *TP53* mutation were not diagnosed with ATC in early age in this study (Table 2-5). However, there might be possible link between these mutations and the aggressive types of thyroid carcinoma since only one out of 579 WDTC patients with germline *TP53* mutation was diagnosed with wiFTC which represents frequent distant metastasis [169]. Thus, it is still not clear how germline *TP53* mutation encourages aggressiveness of thyroid carcinoma, the screening of this gene from patient's genomic DNA would be useful to predict the development of the aggressive thyroid carcinoma.

To sum up the mutational profile results, ATC is developed from PTC or FTC by multiple hits of additional mutations in oncogenes, *TERT*, TSGs, and regulatory genes (Figure 2-24).

Recent two large scale studies broaden the paradigm of classification of WDTC from the histologic to the molecular approach [39]. WDTCs could be categorized into three molecular subtypes; *BRAF*-like, *RAS*-like, and NBNR. *BRAF*-like WDTC is developed by driver genes such as *BRAF*, *BRAF*-fusion, and diverse fusion genes. They were characterized by high activation level of MAPK signaling and cell adhesion molecule pathways as well as down-regulation of metabolic pathways and thyroid differentiation. *H*/*K*/*NRAS* and several fusion genes contributes to the pathogenesis of *RA*S-like WDTC and it exhibits higher thyroid differentiation and less activation of MAPK signaling pathway than *BRAF*-like WDTC. NBNR WDTCs involve alterations in *EIF1AX*, *DICER1*, *IDH1*, and *PAX8-PPARG*. They have highest thyroid differentiation among WDTC and they are not associated with MAPK signaling pathway.

However, unlike WDTC, *BRAF*-positive and *RAS*-positive ATC samples displayed very similar global gene expression and DNA methylation profiles which distinguish them from WDTC (Figure 2-10 and 2-22). In addition, *RAS*-positive ATC samples have even more activation of MAPK signaling

pathway and worse thyroid cell differentiation than *BRAF*-positive ATCs in this study (Figure 11).

Taken together, based on transcriptome and DNA methylation analyses, it could be concluded that ATC exhibits the similar molecular characteristics regardless of gene alteration in contrast to WDTC which represents distinctive molecular profile based on type of altered driver gene.

After successive immunotherapy using pembrolizumab to advanced melanoma, targeting PD-1 and PD-L1 became a promising therapeutic tactic in the field of cancer treatment [209]. PD-1 is a cell surface receptor and expressed on various immune cells [210, 211]. When activated immune cells are abundant in tumor microenvironment, inflammatory cytokines such as INFγ and TNFα induces PD-L1 in tumor cell [212-214]. Interaction between PD-1 and its ligand PD-L1 triggers T-cell anergy and exhaustion, therefore, tumor cells escape from the immune system [215]. Recently, PD-L1 expression in ATC was reported and suggested as a predictive marker [216]. In our analysis, not only PD-L1, but also PD-L2 which is also induces T-cell suppression [217], were up-regulated in *RAS*-positive ATC compared with *RAS*-positive WDTC (Figure 2-16). Furthermore, the up-regulation of *IL6* in ATC was discovered and it was involved in the most significant gene network contributes the progression of FTC origin ATC (Figure 2-14 and 2-15). Anti-IL-6 agents, Siltuximab, Sarilumab, and Tocilizumab, have been demonstrated

as therapeutic target of Castleman disease, rheumatoid arthritis, and even cancer [218-220]. Remarkably, there were recent reports that combination therapy using PD-L1 and IL-6 antibodies on pancreatic and hepatocellular carcinoma efficiently inhibits cancer progression [221, 222].

DNA methylation is the another major strategy for cancer therapy [223]. It is established that CpG island hypermethylaion represses several TSGs, thus it accelerates the progression of cancer. In this study, ATC specific hypermethylation which is enriched in CpG islands was identified and two methyltransferase genes, *DNMT1* and *DNMT3B*, were up-regulated in ATC (Figure 2-23). Reversing abnormal DNA methylation inhibiting these genes have attention in anticancer therapy [224]. However, this result need to be validated further, since it might be derived from the technical artifacts between our *HpaII* digestion WGS and TCGA's microarray data.

Taken together, the immunotherapy using anti-PD-1/PD-L2 blockades together with IL-6-blocking drug as well as DNMT inhibitors could be significant and more efficient therapeutic strategy for ATC.

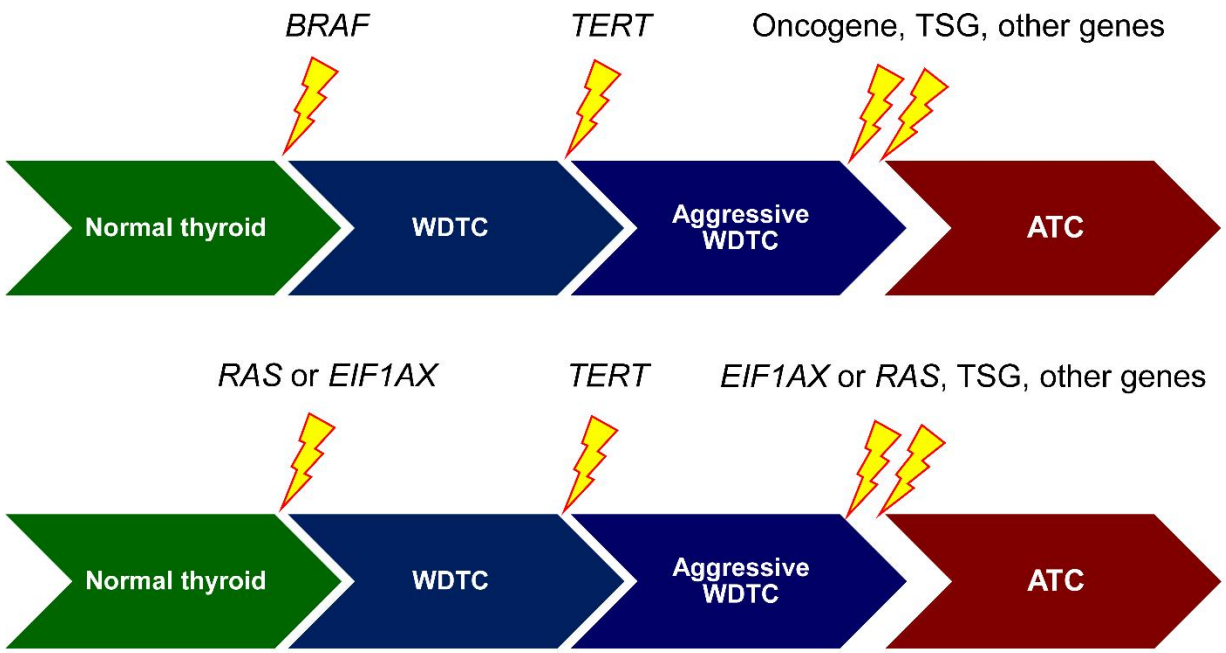

Figure 2-24. A schematic illustration of genomic alterations during thyroid carcinoma progression.

General Conclusion

The objective of this thesis was to decipher the genomic and transcriptomic landscapes of thyroid carcinoma including WDTC and ATC. The major strength of this study is the large and uniform number of diverse subtypes of thyroid carcinoma and integrative analysis of them.

In the first chapter, the molecular characteristics of WDTC including FA, miFTC, FVPTC, and cPTC were illustrated. Until now, the genomic and transcriptomic landscapes of WDTC were described imperfectly, since previous literatures did not analyze the adequate number of papillary- and follicular-patterned thyroid carcinomas at the same time. Based on the results, it could be concluded that WDTC are developed by different types of driver gene alteration which contributes the papillary or follicular histologies. Furthermore, WDTCs are classified as three molecular subtypes (*BRAF*-like, *RAS*-like, and NBNR) and display different dysregulation of signaling pathways such as MAPK, metabolism, ECM receptor, and p53 signaling pathways among them. Interestingly, tumors with *EIF1AX* and *PAX8*-*PPARG* which were defined as *RAS*-like in TCGA study were re-classified as NBNR by the current analysis, hence more expanded analysis on these tumors are required to more stratified therapy. Since these types of tumor are follicular-patterned ones, the re-classification of thyroid tumor based on molecular characteristics is especially important to miFTC, FA, and FVPTC.

In the second chapter, the molecular characteristics of16 ATCs were determined. This is the first study which analyzed the genomic, transcriptomic, and epigenomic landscape of ATC integrates with WDTC. Most of ATC (75%) harbored extra somatic mutations in *TERT*, oncogenes, TSGs, and cancer-related genes in addition to *BRAF* or *H/NRAS*. Moreover, patient with ATC more frequently harbored germline *TP53* mutations than WDTC patients, hence it could be important prognostic indicator. The global gene expression analysis suggested that ATC are distinguishable from WDTCs as the fourth molecular subtype of thyroid carcinoma, ATC-like. Series of analysis revealed dysregulated pathways including MAPK, JAK-STAT, VEGF, and thyroid hormone synthesis signaling in ATC. Furthermore, the potential therapeutic targets such as IL-6, PD-L1, PD-L2, *DNMT1*, and *DNMT3B* were proposed. In particular, the worst thyroid cell differentiation and PD-L1/PD-L2 up-regulation were only identified in ATCs with *CDKN2A*/*CDKN2B* deep deletion. Therefore, it could be suspected that this type of focal deletion might have the significant clinical implication in ATC.

It is expected that the findings in this thesis would lead to more efficient diagnosis, management, and targeted therapeutic strategy against diverse types of thyroid carcinoma.

Reference

1. Yoo, S.K., et al., *Comprehensive Analysis of the Transcriptional and Mutational Landscape of Follicular and Papillary Thyroid Cancers.* PLoS Genet, 2016. **12**(8): p. e1006239.
2. van Dijk, E.L., et al., *Ten years of next-generation sequencing technology.* Trends Genet, 2014. **30**(9): p. 418-26.
3. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors.* Proc Natl Acad Sci U S A, 1977. **74**(12): p. 5463-7.
4. Metzker, M.L., *Sequencing technologies - the next generation.* Nat Rev Genet, 2010. **11**(1): p. 31-46.
5. Eid, J., et al., *Real-time DNA sequencing from single polymerase molecules.* Science, 2009. **323**(5910): p. 133-8.
6. Feng, Y., et al., *Nanopore-based fourth-generation DNA sequencing technology.* Genomics Proteomics Bioinformatics, 2015. **13**(1): p. 4-16.
7. Kim, J.I., et al., *A highly annotated whole-genome sequence of a Korean individual.* Nature, 2009. **460**(7258): p. 1011-5.
8. Sudmant, P.H., et al., *An integrated map of structural variation in 2,504 human genomes.* Nature, 2015. **526**(7571): p. 75-81.
9. Choi, M., et al., *Genetic diagnosis by whole exome capture and massively parallel DNA sequencing.* Proc Natl Acad Sci U S A, 2009. **106**(45): p. 19096-101.

10. Corradin, O. and P.C. Scacheri, *Enhancer variants: evaluating functions in common disease.* Genome Med, 2014. **6**(10): p. 85.

11. Ma, M., et al., *Disease-associated variants in different categories of disease located in distinct regulatory elements.* BMC Genomics, 2015. **16 Suppl 8**: p. S3.

12. Chatterjee, S., et al., *Enhancer Variants Synergistically Drive Dysfunction of a Gene Regulatory Network In Hirschsprung Disease.* Cell, 2016. **167**(2): p. 355-368 e10.

13. Min, B.J., et al., *Whole-exome sequencing identifies mutations of KIF22 in spondyloepimetaphyseal dysplasia with joint laxity, leptodactylic type.* Am J Hum Genet, 2011. **89**(6): p. 760-6.

14. Nuytemans, K., et al., *Whole exome sequencing of rare variants in EIF4G1 and VPS35 in Parkinson disease.* Neurology, 2013. **80**(11): p. 982-9.

15. Lange, L.A., et al., *Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol.* Am J Hum Genet, 2014. **94**(2): p. 233-45.

16. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nat Rev Genet, 2009. **10**(1): p. 57-63.

17. Roberts, A., et al., *Identification of novel transcripts in annotated genomes using RNA-Seq.* Bioinformatics, 2011. **27**(17): p. 2325-9.

18. Ilott, N.E. and C.P. Ponting, *Predicting long non-coding RNAs using RNA sequencing.* Methods, 2013. **63**(1): p. 50-9.

19. Haas, B.J., et al., *De novo transcript sequence reconstruction from*

RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc, 2013. **8**(8): p. 1494-512.

20. Xie, Y., et al., *SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads.* Bioinformatics, 2014. **30**(12): p. 1660-6.
21. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.
22. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.
23. Law, C.W., et al., *voom: Precision weights unlock linear model analysis tools for RNA-seq read counts.* Genome Biol, 2014. **15**(2): p. R29.
24. Lim, B.C., et al., *Hoyeraal-Hreidarsson syndrome with a DKC1 mutation identified by whole-exome sequencing.* Gene, 2014. **546**(2): p. 425-9.
25. Yoo, S.K., et al., *Noninvasive prenatal diagnosis of duchenne muscular dystrophy: comprehensive genetic diagnosis in carrier, proband, and fetus.* Clin Chem, 2015. **61**(6): p. 829-37.
26. Tan, Y., et al., *Clinical outcome of preimplantation genetic diagnosis and screening using next generation sequencing.* Gigascience, 2014. **3**(1): p. 30.
27. Carr, T.H., et al., *Defining actionable mutations for oncology therapeutic development.* Nat Rev Cancer, 2016. **16**(5): p. 319-29.

28. Macconaill, L.E. and L.A. Garraway, *Clinical implications of the cancer genome.* J Clin Oncol, 2010. **28**(35): p. 5219-28.

29. Wooster, R., et al., *Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13.* Science, 1994. **265**(5181): p. 2088-90.

30. Miki, Y., et al., *A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1.* Science, 1994. **266**(5182): p. 66-71.

31. Alaei-Mahabadi, B., et al., *Global analysis of somatic structural genomic alterations and their impact on gene expression in diverse human cancers.* Proc Natl Acad Sci U S A, 2016. **113**(48): p. 13768-13773.

32. Seo, J.S., et al., *The transcriptional landscape and mutational profile of lung adenocarcinoma.* Genome Res, 2012. **22**(11): p. 2109-19.

33. Ju, Y.S., et al., *A transforming KIF5B and RET gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing.* Genome Res, 2012. **22**(3): p. 436-45.

34. Min, B.H., et al., *Dysregulated Wnt signalling and recurrent mutations of the tumour suppressor RNF43 in early gastric carcinogenesis.* J Pathol, 2016. **240**(3): p. 304-314.

35. Cancer Genome Atlas Research, N., *Comprehensive genomic characterization defines human glioblastoma genes and core pathways.* Nature, 2008. **455**(7216): p. 1061-8.

36. Cancer Genome Atlas Research, N., *Integrated genomic analyses of ovarian carcinoma.* Nature, 2011. **474**(7353): p. 609-15.

37. Cancer Genome Atlas, N., *Comprehensive molecular characterization of human colon and rectal cancer.* Nature, 2012. **487**(7407): p. 330-7.

38. Cancer Genome Atlas Research, N., *Comprehensive molecular characterization of clear cell renal cell carcinoma.* Nature, 2013. **499**(7456): p. 43-9.

39. Cancer Genome Atlas Research, N., *Integrated genomic characterization of papillary thyroid carcinoma.* Cell, 2014. **159**(3): p. 676-90.

40. Zheng, S., et al., *Comprehensive Pan-Genomic Characterization of Adrenocortical Carcinoma.* Cancer Cell, 2016. **29**(5): p. 723-36.

41. Alexandrov, L.B., et al., *Signatures of mutational processes in human cancer.* Nature, 2013. **500**(7463): p. 415-21.

42. Stratton, M.R., P.J. Campbell, and P.A. Futreal, *The cancer genome.* Nature, 2009. **458**(7239): p. 719-24.

43. Liu, H., et al., *Remarkable difference of somatic mutation patterns between oncogenes and tumor suppressor genes.* Oncol Rep, 2011. **26**(6): p. 1539-46.

44. Cibulskis, K., et al., *Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples.* Nat Biotechnol, 2013. **31**(3): p. 213-9.

45. Shaw, A.T., et al., *Resensitization to Crizotinib by the Lorlatinib ALK Resistance Mutation L1198F.* N Engl J Med, 2016. **374**(1): p. 54-61.

46. Li, J., et al., *L1198F Mutation Resensitizes Crizotinib to ALK by Altering the Conformation of Inhibitor and ATP Binding Sites.* Int J

Mol Sci, 2017. **18**(3).

47. Oxnard, G.R., A. Binder, and P.A. Janne, *New targetable oncogenes in non-small-cell lung cancer.* J Clin Oncol, 2013. **31**(8): p. 1097-104.

48. Helleday, T., S. Eshtad, and S. Nik-Zainal, *Mechanisms underlying mutational signatures in human cancers.* Nat Rev Genet, 2014. **15**(9): p. 585-98.

49. D'Antonio, M., et al., *Kataegis Expression Signature in Breast Cancer Is Associated with Late Onset, Better Prognosis, and Higher HER2 Levels.* Cell Rep, 2016. **16**(3): p. 672-83.

50. Robertson, A.G., et al., *Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer.* Cell, 2017. **171**(3): p. 540-556 e25.

51. Aplan, P.D., *Causes of oncogenic chromosomal translocation.* Trends Genet, 2006. **22**(1): p. 46-55.

52. Parker, B.C., et al., *The tumorigenic FGFR3-TACC3 gene fusion escapes miR-99a regulation in glioblastoma.* J Clin Invest, 2013. **123**(2): p. 855-65.

53. Ji, J.H., et al., *Identification of Driving ALK Fusion Genes and Genomic Landscape of Medullary Thyroid Cancer.* PLoS Genet, 2015. **11**(8): p. e1005467.

54. Stransky, N., et al., *The landscape of kinase fusions in cancer.* Nat Commun, 2014. **5**: p. 4846.

55. Chan, B.A. and B.G. Hughes, *Targeted therapy for non-small cell lung cancer: current standards and the promise of the future.* Transl Lung Cancer Res, 2015. **4**(1): p. 36-54.

56. Li, Y.Y., et al., *Exome and genome sequencing of nasopharynx cancer identifies NF-kappaB pathway activating mutations.* Nat Commun, 2017. **8**: p. 14121.

57. Beroukhim, R., et al., *The landscape of somatic copy-number alteration across human cancers.* Nature, 2010. **463**(7283): p. 899-905.

58. Alhejaily, A., et al., *Inactivation of the CDKN2A tumor-suppressor gene by deletion or methylation is common at diagnosis in follicular lymphoma and associated with poor clinical outcome.* Clin Cancer Res, 2014. **20**(6): p. 1676-86.

59. Kaneshiro, D., et al., *Chromosome 1p and 19q deletions in glioblastoma multiforme.* Appl Immunohistochem Mol Morphol, 2009. **17**(6): p. 512-6.

60. Cordoba, I., et al., *Better prognosis for patients with del(7q) than for patients with monosomy 7 in myelodysplastic syndrome.* Cancer, 2012. **118**(1): p. 127-33.

61. Zare, F., et al., *An evaluation of copy number variation detection tools for cancer using whole exome sequencing data.* BMC Bioinformatics, 2017. **18**(1): p. 286.

62. Olshen, A.B., et al., *Circular binary segmentation for the analysis of array-based DNA copy number data.* Biostatistics, 2004. **5**(4): p. 557-72.

63. Kuilman, T., et al., *CopywriteR: DNA copy number detection from off-target sequence data.* Genome Biol, 2015. **16**: p. 49.

64. D'Aurizio, R., et al., *Enhanced copy number variants detection from whole-exome sequencing data using EXCAVATOR2*. Nucleic Acids Res, 2016. **44**(20): p. e154.

65. Guo, X., et al., *Clinically Relevant Molecular Subtypes in Leiomyosarcoma.* Clin Cancer Res, 2015. **21**(15): p. 3501-11.

66. Woo, H.G., et al., *Integrative analysis of genomic and epigenomic regulation of the transcriptome in liver cancer.* Nat Commun, 2017. **8**(1): p. 839.

67. Toss, A. and M. Cristofanilli, *Molecular characterization and targeted therapeutic approaches in breast cancer.* Breast Cancer Res, 2015. **17**: p. 60.

68. Graham, D.M., et al., *Molecular Subtypes and Personalized Therapy in Metastatic Colorectal Cancer.* Curr Colorectal Cancer Rep, 2016. **12**: p. 141-150.

69. Xing, M., *Molecular pathogenesis and mechanisms of thyroid cancer.* Nat Rev Cancer, 2013. **13**(3): p. 184-99.

70. Rahbari, R., L. Zhang, and E. Kebebew, *Thyroid cancer gender disparity.* Future Oncol, 2010. **6**(11): p. 1771-9.

71. Kitahara, C.M. and J.A. Sosa, *The changing incidence of thyroid cancer.* Nat Rev Endocrinol, 2016. **12**(11): p. 646-653.

72. Oh, C.M., et al., *Cancer Statistics in Korea: Incidence, Mortality, Survival, and Prevalence in 2013.* Cancer Res Treat, 2016. **48**(2): p. 436-50.

73. Jegerlehner, S., et al., *Overdiagnosis and overtreatment of thyroid*

cancer: A population-based temporal trend study. PLoS One, 2017. **12**(6): p. e0179387.

74. La Vecchia, C., et al., *Thyroid cancer mortality and incidence: a global overview.* Int J Cancer, 2015. **136**(9): p. 2187-95.

75. Schneider, D.F. and H. Chen, *New developments in the diagnosis and treatment of thyroid cancer.* CA Cancer J Clin, 2013. **63**(6): p. 374-94.

76. Paschke, R., et al., *The Treatment of Well-Differentiated Thyroid Carcinoma.* Dtsch Arztebl Int, 2015. **112**(26): p. 452-8.

77. Sippel, R.S., M. Kunnimalaiyaan, and H. Chen, *Current management of medullary thyroid cancer.* Oncologist, 2008. **13**(5): p. 539-47.

78. Luster, M., et al., *Radioiodine therapy of metastatic lesions of differentiated thyroid cancer.* J Endocrinol Invest, 2012. **35**(6 Suppl): p. 21-9.

79. Valerio, L., et al., *Targeted Therapy in Thyroid Cancer: State of the Art.* Clin Oncol (R Coll Radiol), 2017. **29**(5): p. 316-324.

80. Durante, C., et al., *Long-term outcome of 444 patients with distant metastases from papillary and follicular thyroid carcinoma: benefits and limits of radioiodine therapy.* J Clin Endocrinol Metab, 2006. **91**(8): p. 2892-9.

81. Costa, V., et al., *New somatic mutations and WNK1-B4GALNT3 gene fusion in papillary thyroid carcinoma.* Oncotarget, 2015. **6**(13): p. 11242-51.

82. Siraj, A.K., et al., *Genomic Profiling of Thyroid Cancer Reveals a Role for Thyroglobulin in Metastasis.* Am J Hum Genet, 2016. **98**(6):

p. 1170-1180.

83. Jung, S.H., et al., *Mutational burdens and evolutionary ages of thyroid follicular adenoma are comparable to those of follicular carcinoma.* Oncotarget, 2016. **7**(43): p. 69638-69648.

84. McHenry, C.R. and R. Phitayakorn, *Follicular adenoma and carcinoma of the thyroid gland.* Oncologist, 2011. **16**(5): p. 585-93.

85. Keutgen, X.M., S.M. Sadowski, and E. Kebebew, *Management of anaplastic thyroid cancer.* Gland Surg, 2015. **4**(1): p. 44-51.

86. Hundahl, S.A., et al., *A National Cancer Data Base report on 53,856 cases of thyroid carcinoma treated in the U.S., 1985-1995 [see commetns]*. Cancer, 1998. **83**(12): p. 2638-48.

87. Kebebew, E., et al., *Anaplastic thyroid carcinoma. Treatment outcome and prognostic factors*. Cancer, 2005. **103**(7): p. 1330-5.

88. Haigh, P.I., et al., *Completely resected anaplastic thyroid carcinoma combined with adjuvant chemotherapy and irradiation is associated with prolonged survival.* Cancer, 2001. **91**(12): p. 2335-42.

89. Molinaro, E., et al., *Anaplastic thyroid carcinoma: from clinicopathology to genetics and advanced therapies.* Nat Rev Endocrinol, 2017. **13**(11): p. 644-660.

90. Are, C. and A.R. Shaha, *Anaplastic thyroid carcinoma: biology, pathogenesis, prognostic factors, and treatment approaches.* Ann Surg Oncol, 2006. **13**(4): p. 453-64.

91. Bronner, M.P. and V.A. LiVolsi, *Spindle cell squamous carcinoma of the thyroid: an unusual anaplastic tumor associated with tall cell*

papillary cancer. Mod Pathol, 1991. **4**(5): p. 637-43.

92. Nikiforova, M.N. and Y.E. Nikiforov, *Molecular diagnostics and predictors in thyroid cancer.* Thyroid, 2009. **19**(12): p. 1351-61.

93. Derbel, O., et al., *Results of combined treatment of anaplastic thyroid carcinoma (ATC).* BMC Cancer, 2011. **11**: p. 469.

94. He, X., et al., *Outcome after intensity modulated radiotherapy for anaplastic thyroid carcinoma.* BMC Cancer, 2014. **14**: p. 235.

95. Nagaiah, G., et al., *Anaplastic thyroid cancer: a review of epidemiology, pathogenesis, and treatment.* J Oncol, 2011. **2011**: p. 542358.

96. Besic, N., et al., *Prognostic factors in anaplastic carcinoma of the thyroid-a multivariate survival analysis of 188 patients.* Langenbecks Arch Surg, 2005. **390**(3): p. 203-8.

97. Besic, N. and B. Gazic, *Sites of metastases of anaplastic thyroid carcinoma: autopsy findings in 45 cases from a single institution.* Thyroid, 2013. **23**(6): p. 709-13.

98. Jereb, B., J. Stjernsward, and T. Lowhagen, *Anaplastic giant-cell carcinoma of the thyroid. A study of treatment and prognosis.* Cancer, 1975. **35**(5): p. 1293-5.

99. Kunstman, J.W., et al., *Characterization of the mutational landscape of anaplastic thyroid cancer via whole-exome sequencing.* Hum Mol Genet, 2015. **24**(8): p. 2318-29.

100. Kasaian, K., et al., *The genomic and transcriptomic landscape of anaplastic thyroid cancer: implications for therapy.* BMC Cancer,

2015. **15**: p. 984.

101. Landa, I., et al., *Genomic and transcriptomic hallmarks of poorly differentiated and anaplastic thyroid cancers.* J Clin Invest, 2016. **126**(3): p. 1052-66.

102. Zhao, S., et al., *Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells.* PLoS One, 2014. **9**(1): p. e78644.

103. Jeon, M.J., et al., *Genomic Alterations of Anaplastic Thyroid Carcinoma Detected by Targeted Massive Parallel Sequencing in a BRAF(V600E) Mutation-Prevalent Area.* Thyroid, 2016. **26**(5): p. 683-90.

104. Bonhomme, B., et al., *Molecular Pathology of Anaplastic Thyroid Carcinomas: A Retrospective Study of 144 Cases.* Thyroid, 2017. **27**(5): p. 682-692.

105. Tiedje, V., et al., *NGS based identification of mutational hotspots for targeted therapy in anaplastic thyroid carcinoma.* Oncotarget, 2017. **8**(26): p. 42613-42620.

106. Schmid, K.W. and N.R. Farid, *How to define follicular thyroid carcinoma?* Virchows Arch, 2006. **448**(4): p. 385-93.

107. Asa, S.L., T.J. Giordano, and V.A. LiVolsi, *Implications of the TCGA genomic characterization of papillary thyroid carcinoma for thyroid pathology: does follicular variant papillary thyroid carcinoma exist?* Thyroid, 2015. **25**(1): p. 1-2.

108. Vasko, V., et al., *Specific pattern of RAS oncogene mutations in*

follicular thyroid tumors. J Clin Endocrinol Metab, 2003. **88**(6): p. 2745-52.

109. Zhu, Z., et al., *Molecular profile and clinical-pathologic features of the follicular variant of papillary thyroid carcinoma. An unusually high prevalence of ras mutations.* Am J Clin Pathol, 2003. **120**(1): p. 71-7.

110. Fukahori, M., et al., *The associations between RAS mutations and clinical characteristics in follicular thyroid tumors: new insights from a single center and a large patient cohort.* Thyroid, 2012. **22**(7): p. 683-9.

111. Nikiforov, Y.E., P.W. Biddinger, and L.D.R. Thompson, *Diagnostic Pathology and Molecular Genetics of the Thyroid: A Comprehensive Guide for Practicing Thyroid Pathology*. 2012: Wolters Kluwer Health.

112. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.

113. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Res, 2010. **20**(9): p. 1297-303.

114. Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.* Nucleic Acids Res, 2010. **38**(16): p. e164.

115. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans.* Nature, 2016. **536**(7616): p. 285-91.

116. Cerami, E., et al., *The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.* Cancer Discov, 2012. **2**(5): p. 401-4.

117. Anders, S., P.T. Pyl, and W. Huber, *HTSeq--a Python framework to work with high-throughput sequencing data.* Bioinformatics, 2015. **31**(2): p. 166-9.

118. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.* Journal of the Royal Statistical Society. Series B (Methodological), 1995. **57**(1): p. 289-300.

119. de Hoon, M.J., et al., *Open source clustering software.* Bioinformatics, 2004. **20**(9): p. 1453-4.

120. Liberzon, A., et al., *Molecular signatures database (MSigDB) 3.0.* Bioinformatics, 2011. **27**(12): p. 1739-40.

121. Pratilas, C.A., et al., *(V600E)BRAF is associated with disabled feedback inhibition of RAF-MEK signaling and elevated transcriptional output of the pathway.* Proc Natl Acad Sci U S A, 2009. **106**(11): p. 4519-24.

122. Heravi-Moussavi, A., et al., *Recurrent somatic DICER1 mutations in nonepithelial ovarian cancers.* N Engl J Med, 2012. **366**(3): p. 234-42.

123. de Kock, L., et al., *Exploring the association Between DICER1 mutations and differentiated thyroid carcinoma.* J Clin Endocrinol Metab, 2014. **99**(6): p. E1072-7.

124. Karunamurthy, A., et al., *Prevalence and phenotypic correlations of*

EIF1AX mutations in thyroid nodules. Endocr Relat Cancer, 2016. **23**(4): p. 295-301.

125. Chang, M.T., et al., *Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity.* Nat Biotechnol, 2016. **34**(2): p. 155-63.

126. Ye, L., et al., *The genetic landscape of benign thyroid nodules revealed by whole exome and transcriptome sequencing.* Nat Commun, 2017. **8**: p. 15533.

127. Le Gallo, M., et al., *Exome sequencing of serous endometrial tumors identifies recurrent somatic mutations in chromatin-remodeling and ubiquitin ligase complex genes.* Nat Genet, 2012. **44**(12): p. 1310-5.

128. Mani, R.S., *The emerging role of speckle-type POZ protein (SPOP) in cancer development.* Drug Discov Today, 2014. **19**(9): p. 1498-502.

129. Calebiro, D., et al., *Recurrent EZH1 mutations are a second hit in autonomous thyroid adenomas.* J Clin Invest, 2016. **126**(9): p. 3383-8.

130. Guldberg, P., et al., *Somatic mutation of the Peutz-Jeghers syndrome gene, LKB1/STK11, in malignant melanoma.* Oncogene, 1999. **18**(9): p. 1777-80.

131. Su, G.H., et al., *Germline and somatic mutations of the STK11/LKB1 Peutz-Jeghers gene in pancreatic and biliary cancers.* Am J Pathol, 1999. **154**(6): p. 1835-40.

132. Zhou, J.B., et al., *Identification of a novel gene fusion RNF213SLC26A11 in chronic myeloid leukemia by RNA-Seq.* Mol Med Rep, 2013. **7**(2): p. 591-7.

133. Nakaoku, T., et al., *Druggable oncogene fusions in invasive mucinous lung adenocarcinoma.* Clin Cancer Res, 2014. **20**(12): p. 3087-93.

134. Ross, J.S., et al., *New routes to targeted therapy of intrahepatic cholangiocarcinomas revealed by next-generation sequencing.* Oncologist, 2014. **19**(3): p. 235-42.

135. Panebianco, F., et al., *THADA fusion is a mechanism of IGF2BP3 activation and IGF1R signaling in thyroid cancer.* Proc Natl Acad Sci U S A, 2017. **114**(9): p. 2307-2312.

136. Montagut, C. and J. Settleman, *Targeting the RAF-MEK-ERK pathway in cancer therapy.* Cancer Lett, 2009. **283**(2): p. 125-34.

137. Nikiforov, Y.E., et al., *Nomenclature Revision for Encapsulated Follicular Variant of Papillary Thyroid Carcinoma: A Paradigm Shift to Reduce Overtreatment of Indolent Tumors.* JAMA Oncol, 2016. **2**(8): p. 1023-9.

138. Maximo, V. and M. Sobrinho-Simoes, *Hurthle cell tumours of the thyroid. A review with emphasis on mitochondrial abnormalities with clinical relevance.* Virchows Arch, 2000. **437**(2): p. 107-15.

139. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic Acids Res, 2000. **28**(1): p. 27-30.

140. Chudova, D., et al., *Molecular classification of thyroid nodules using high-dimensionality genomic data.* J Clin Endocrinol Metab, 2010. **95**(12): p. 5296-304.

141. Nucera, C., J. Lawler, and S. Parangi, *BRAF(V600E) and microenvironment in thyroid cancer: a functional link to drive cancer*

progression. Cancer Res, 2011. **71**(7): p. 2417-22.

142. Shan, Z., et al., *Gadd45a inhibits cell migration and invasion by altering the global RNA expression.* Cancer Biol Ther, 2012. **13**(11): p. 1112-22.

143. Davis, C.F., et al., *The somatic genomic landscape of chromophobe renal cell carcinoma.* Cancer Cell, 2014. **26**(3): p. 319-330.

144. Wreesmann, V.B., et al., *Follicular variant of papillary thyroid carcinoma: genome-wide appraisal of a controversial entity.* Genes Chromosomes Cancer, 2004. **40**(4): p. 355-64.

145. Melillo, R.M., et al., *The RET/PTC-RAS-BRAF linear signaling cascade mediates the motile and mitogenic phenotype of thyroid cancer cells.* J Clin Invest, 2005. **115**(4): p. 1068-81.

146. Wu, Z., et al., *Mechanisms controlling mitochondrial biogenesis and respiration through the thermogenic coactivator PGC-1.* Cell, 1999. **98**(1): p. 115-24.

147. Schreiber, S.N., et al., *The estrogen-related receptor alpha (ERRalpha) functions in PPARgamma coactivator 1alpha (PGC-1alpha)-induced mitochondrial biogenesis.* Proc Natl Acad Sci U S A, 2004. **101**(17): p. 6472-7.

148. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.

149. Rubinstein, J.C., et al., *Shifting patterns of genomic variation in the somatic evolution of papillary thyroid carcinoma.* BMC Cancer, 2016.

16: p. 646.

150. Leinonen, R., et al., *The sequence read archive.* Nucleic Acids Res, 2011. **39**(Database issue): p. D19-21.

151. Saunders, C.T., et al., *Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs.* Bioinformatics, 2012. **28**(14): p. 1811-7.

152. Genomes Project, C., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.

153. Gehring, J.S., et al., *SomaticSignatures: inferring mutational signatures from single-nucleotide variants.* Bioinformatics, 2015. **31**(22): p. 3673-5.

154. Shen, R. and V.E. Seshan, *FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing.* Nucleic Acids Res, 2016. **44**(16): p. e131.

155. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.

156. Mermel, C.H., et al., *GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers.* Genome Biol, 2011. **12**(4): p. R41.

157. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis.* BMC Bioinformatics, 2008. **9**: p. 559.

158. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.

159. Reyna-Lopez, G.E., J. Simpson, and J. Ruiz-Herrera, *Differences in DNA methylation patterns are detectable during the dimorphic transition of fungi by amplification of restriction polymorphisms.* Mol Gen Genet, 1997. **253**(6): p. 703-10.

160. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* Bioinformatics, 2010. **26**(6): p. 841-2.

161. Cancer Genome Atlas Research, N., *Comprehensive genomic characterization of squamous cell lung cancers.* Nature, 2012. **489**(7417): p. 519-25.

162. Li, C., et al., *Whole Exome Sequencing Identifies Frequent Somatic Mutations in Cell-Cell Adhesion Genes in Chinese Patients with Lung Squamous Cell Carcinoma.* Sci Rep, 2015. **5**: p. 14237.

163. Muller, P.A. and K.H. Vousden, *Mutant p53 in cancer: new functions and therapeutic opportunities.* Cancer Cell, 2014. **25**(3): p. 304-17.

164. Cimmino, L., et al., *TET1 is a tumor suppressor of hematopoietic malignancy.* Nat Immunol, 2015. **16**(6): p. 653-62.

165. Neri, F., et al., *TET1 is a tumour suppressor that inhibits colon cancer growth by derepressing inhibitors of the WNT pathway.* Oncogene, 2015. **34**(32): p. 4168-76.

166. Rao, R.C. and Y. Dou, *Hijacked in cancer: the KMT2 (MLL) family of methyltransferases.* Nat Rev Cancer, 2015. **15**(6): p. 334-46.

167. Napier, C.E., et al., *ATRX represses alternative lengthening of telomeres.* Oncotarget, 2015. **6**(18): p. 16543-58.

168. Lovejoy, C.A., et al., *Loss of ATRX, genome instability, and an altered*

DNA damage response are hallmarks of the alternative lengthening of telomeres pathway. PLoS Genet, 2012. **8**(7): p. e1002772.

169. Podda, M., et al., *Follicular thyroid carcinoma: differences in clinical relevance between minimally invasive and widely invasive tumors.* World J Surg Oncol, 2015. **13**: p. 193.

170. Siraj, A.K., et al., *ALK alteration is a frequent event in aggressive breast cancers.* Breast Cancer Res, 2015. **17**: p. 127.

171. Kim, M.S., et al., *Mutational analysis of oncogenic AKT E17K mutation in common solid cancers and acute leukaemias.* Br J Cancer, 2008. **98**(9): p. 1533-5.

172. Carpten, J.D., et al., *A transforming mutation in the pleckstrin homology domain of AKT1 in cancer.* Nature, 2007. **448**(7152): p. 439-44.

173. Karakas, B., K.E. Bachman, and B.H. Park, *Mutation of the PIK3CA oncogene in human cancers.* Br J Cancer, 2006. **94**(4): p. 455-9.

174. Samuels, Y., et al., *Mutant PIK3CA promotes cell growth and invasion of human cancer cells.* Cancer Cell, 2005. **7**(6): p. 561-73.

175. Gibson, W.J., et al., *Genomic Heterogeneity and Exceptional Response to Dual Pathway Inhibition in Anaplastic Thyroid Cancer.* Clin Cancer Res, 2017. **23**(9): p. 2367-2373.

176. Yu, T., J. Bachman, and Z.C. Lai, *Evidence for a tumor suppressor role for the large tumor suppressor genes LATS1 and LATS2 in human cancer.* Genetics, 2013. **195**(3): p. 1193-6.

177. Yu, T., J. Bachman, and Z.C. Lai, *Mutation analysis of large tumor*

suppressor genes LATS1 and LATS2 supports a tumor suppressor role in human cancer. Protein Cell, 2015. **6**(1): p. 6-11.

178. Graubert, T.A., et al., *Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes.* Nat Genet, 2011. **44**(1): p. 53-7.

179. Przychodzen, B., et al., *Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms.* Blood, 2013. **122**(6): p. 999-1006.

180. Imielinski, M., et al., *Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing.* Cell, 2012. **150**(6): p. 1107-20.

181. Conticello, S.G., *The AID/APOBEC family of nucleic acid mutators.* Genome Biol, 2008. **9**(6): p. 229.

182. Nabel, C.S., et al., *AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation.* Nat Chem Biol, 2012. **8**(9): p. 751-8.

183. Sieuwerts, A.M., et al., *Elevated APOBEC3B correlates with poor outcomes for estrogen-receptor-positive breast cancers.* Horm Cancer, 2014. **5**(6): p. 405-13.

184. Cescon, D.W., B. Haibe-Kains, and T.W. Mak, *APOBEC3B expression in breast cancer reflects cellular proliferation, while a deletion polymorphism is associated with immune activation.* Proc Natl Acad Sci U S A, 2015. **112**(9): p. 2841-6.

185. Yan, S., et al., *Increased APOBEC3B Predicts Worse Outcomes in Lung Cancer: A Comprehensive Retrospective Study.* J Cancer, 2016. **7**(6): p. 618-25.

186. Borras, E., et al., *Genomic Landscape of Colorectal Mucosa and*

Adenomas. Cancer Prev Res (Phila), 2016. **9**(6): p. 417-27.

187. Shlien, A., et al., *Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers.* Nat Genet, 2015. **47**(3): p. 257-62.

188. Miller, A., et al., *High somatic mutation and neoantigen burden are correlated with decreased progression-free survival in multiple myeloma.* Blood Cancer J, 2017. **7**(9): p. e612.

189. Barretina, J., et al., *The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity.* Nature, 2012. **483**(7391): p. 603-7.

190. Pasello, G., et al., *DNA copy number alterations correlate with survival of esophageal adenocarcinoma patients.* Mod Pathol, 2009. **22**(1): p. 58-65.

191. Pinto, A.E., et al., *DNA ploidy is an independent predictor of survival in breast invasive ductal carcinoma: a long-term multivariate analysis of 393 patients.* Ann Surg Oncol, 2013. **20**(5): p. 1530-7.

192. Carter, S.L., et al., *A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers.* Nat Genet, 2006. **38**(9): p. 1043-8.

193. Jourdan, M., et al., *A major role for Mcl-1 antiapoptotic protein in the IL-6-induced survival of human myeloma cells.* Oncogene, 2003. **22**(19): p. 2950-9.

194. Shi, Y., et al., *IL-6-induced stimulation of c-myc translation in multiple myeloma cells is mediated by myc internal ribosome entry*

site function and the RNA-binding protein, hnRNP A1. Cancer Res, 2008. **68**(24): p. 10215-22.

195. Zitvogel, L. and G. Kroemer, *Targeting PD-1/PD-L1 interactions for cancer immunotherapy.* Oncoimmunology, 2012. **1**(8): p. 1223-1225.

196. Yearley, J.H., et al., *PD-L2 Expression in Human Tumors: Relevance to Anti-PD-1 Therapy in Cancer.* Clin Cancer Res, 2017. **23**(12): p. 3158-3167.

197. Clarke, J.M. and H.I. Hurwitz, *Targeted inhibition of VEGF receptor 2: an update on ramucirumab.* Expert Opin Biol Ther, 2013. **13**(8): p. 1187-96.

198. Robertson, K.D., *DNA methylation, methyltransferases, and cancer.* Oncogene, 2001. **20**(24): p. 3139-55.

199. Beaulieu, N., et al., *An essential role for DNA methyltransferase DNMT3B in cancer cell survival.* J Biol Chem, 2002. **277**(31): p. 28176-81.

200. Wu, C.T., et al., *Expression and function role of DNA methyltransferase 1 in human bladder cancer.* Cancer, 2011. **117**(22): p. 5221-33.

201. Valentijn, L.J., et al., *TERT rearrangements are frequent in neuroblastoma and identify aggressive tumors.* Nat Genet, 2015. **47**(12): p. 1411-4.

202. Peifer, M., et al., *Telomerase activation by genomic rearrangements in high-risk neuroblastoma.* Nature, 2015. **526**(7575): p. 700-4.

203. Horn, S., et al., *TERT promoter mutations in familial and sporadic*

melanoma. Science, 2013. **339**(6122): p. 959-61.

204. Liu, X., et al., *Highly prevalent TERT promoter mutations in aggressive thyroid cancers.* Endocr Relat Cancer, 2013. **20**(4): p. 603-10.

205. Melo, M., et al., *TERT promoter mutations are a major indicator of poor outcome in differentiated thyroid carcinomas.* J Clin Endocrinol Metab, 2014. **99**(5): p. E754-65.

206. Qu, Y., et al., *TERT promoter mutations predict worse survival in laryngeal cancer patients.* Int J Cancer, 2014. **135**(4): p. 1008-10.

207. Nichols, K.E., et al., *Germ-line p53 mutations predispose to a wide spectrum of early-onset cancers.* Cancer Epidemiol Biomarkers Prev, 2001. **10**(2): p. 83-7.

208. Yurgelun, M.B., et al., *Germline TP53 Mutations in Patients With Early-Onset Colorectal Cancer in the Colon Cancer Family Registry.* JAMA Oncol, 2015. **1**(2): p. 214-21.

209. Robert, C., et al., *Pembrolizumab versus Ipilimumab in Advanced Melanoma.* N Engl J Med, 2015. **372**(26): p. 2521-32.

210. Riella, L.V., et al., *Role of the PD-1 pathway in the immune response.* Am J Transplant, 2012. **12**(10): p. 2575-87.

211. Riley, J.L., *PD-1 signaling in primary T cells.* Immunol Rev, 2009. **229**(1): p. 114-25.

212. Ou, J.N., A.E. Wiedeman, and A.M. Stevens, *TNF-alpha and TGF-beta counter-regulate PD-L1 expression on monocytes in systemic lupus erythematosus.* Sci Rep, 2012. **2**: p. 295.

213. Abiko, K., et al., *IFN-gamma from lymphocytes induces PD-L1 expression and promotes progression of ovarian cancer.* Br J Cancer, 2015. **112**(9): p. 1501-9.

214. Mandai, M., et al., *Dual Faces of IFNgamma in Cancer Progression: A Role of PD-L1 Induction in the Determination of Pro- and Antitumor Immunity.* Clin Cancer Res, 2016. **22**(10): p. 2329-34.

215. Chen, L. and X. Han, *Anti-PD-1/PD-L1 therapy of human cancer: past, present, and future.* J Clin Invest, 2015. **125**(9): p. 3384-91.

216. Chintakuntlawar, A.V., et al., *Expression of PD-1 and PD-L1 in anaplastic thyroid cancer patients treated with multimodal therapy: results from a retrospective study.* J Clin Endocrinol Metab, 2017.

217. Rozali, E.N., et al., *Programmed death ligand 2 in cancer-induced immune suppression.* Clin Dev Immunol, 2012. **2012**: p. 656340.

218. van Rhee, F., et al., *Siltuximab for multicentric Castleman's disease: a randomised, double-blind, placebo-controlled trial.* Lancet Oncol, 2014. **15**(9): p. 966-74.

219. Korneev, K.V., et al., *TLR-signaling and proinflammatory cytokines as drivers of tumorigenesis.* Cytokine, 2017. **89**: p. 127-135.

220. Lee, E.B., *A review of sarilumab for the treatment of rheumatoid arthritis.* Immunotherapy, 2017.

221. Liu, H., J. Shen, and K. Lu, *IL-6 and PD-L1 blockade combination inhibits hepatocellular carcinoma cancer development in mouse model.* Biochem Biophys Res Commun, 2017. **486**(2): p. 239-244.

222. Mace, T.A., et al., *IL-6 and PD-L1 antibody blockade combination*

therapy reduces tumour progression in murine models of pancreatic cancer. Gut, 2016.

223. Issa, J.P., *DNA methylation as a therapeutic target in cancer*. Clin Cancer Res, 2007. **13**(6): p. 1634-7.

224. Gnyszka, A., Z. Jastrzebski, and S. Flis, *DNA methyltransferase inhibitors and their emerging role in epigenetic therapy of cancer*. Anticancer Res, 2013. **33**(8): p. 2989-96.

국문초록

차세대 염기서열 분석을 활용한 분화갑상선종양 180 례와 미분화갑상선암 16 례의 유전체 및 전사체 분석 연구

서울대학교 자연과학대학원 협동과정 생물정보학

유 승 근

갑상선암은 가장 흔한 내분비계의 암 중 하나로, 분화갑상선암(well-differentiated thyroid carcinoma)이 90%를 차지하며 대부분 갑상선유두암 (papillary thyroid carcinoma)과 갑상선여포암(follicular thyroid carcinoma)으로 분류된다. 최근, 차세대 염기서열 분석 방법의 발전에 힘입어 분화갑상선암에 대한 분자유전학적 특성이 활발히 연구 되었지만, 해당 연구들은 대부분 갑상선유두암에 집중되어 이루어진 실정이다. 갑상선여포선종(follicular adenoma)은 갑상선여포암과 세포학적 특성이 매우 유사하여 수술 전 검사로 감별할 수 없기 때문에, 수술 후 조직병리학적 현미경 검사를 통해서만 최종 감별진단이 가능하다. 이로 인해 갑상선여포암과 갑상선여포선종은 갑상선여포성종양 (follicular

thyroid neoplasm)으로 통칭되고 있으며 보다 정확하고 효율적인 감별진단을 위한 바이오마커가 필요한 실정이다. 미분화갑상선암 (anaplastic thyroid carcinoma)은 대표적인 희귀 난치성 암의 하나로서, 매우 희귀하지만 전체 갑상선암으로 의한 사망 중 50%를 차지한다. 하지만 이러한 임상적 중요성에도 불구하고, 미분화갑상선암의 발달에 대한 대규모 분자유전학적 특성에 대한 연구는 활발히 이루어 지지 않은 실정이다. 본 연구에서는 현재까지 밝혀지지 않은 갑상선여포성종양과 미분화갑상선암에 대한 분자유전학적 특성을 차세대 염기서열 분석 기법을 통해 알아보고자 하였다. 첫번째 연구에서는 전사체 시퀀싱 기법을 활용하여 30 개의 미세침습 여포성갑상선암 (minimally invasive follicular thyroid carcinoma), 25 개의 갑상선여포선종과 함께 125 개의 갑상선유두암에 대해 분석을 실시하였다. 갑상선여포성종양은 *H/K/NRAS*, *DICER1*, *EIF1AX*, *EZH1*, *IDH1*, *PTEN*, *SOS1*, *SPOP* 등의 유전자들에 생기는 점 돌연변이에 의해 주로 발달되었으며 *PAX8-PPARG* 만이 융합유전자로서 발굴되었다. 반면에, 갑상선유두암에는 *BRAF* 와 다양한 종류의 융합유전자 (*ETV6-NTRK3, NCOA4-RET, CCDC6-RET* 등)가 호발하였다. 이러한 유전자들의 변이 종류에 따라 분화갑상선암은 3 가지 분자적 아형인 *BRAF*-like, *RAS*-like, Non-*BRAF*-Non-*RAS* (NBNR)으로 세포학적 특성과 관계없이 구분될 수 있었다. 또한, 미세침습

여포성갑상선암과 갑상선여포선종은 전사체 수준에서 세포학적 특성과 마찬가지로 구분되지 않아, 갑상선여포성종양에 대한 새로운 진단 및 치료 기준의 필요성을 제시하였다. 두번째 연구에서는 16 개의 미분화갑상선암 조직에 대하여 전장 유전체 서열 및 전사체 분석을 수행하였다. 이를 통하여 갑상선암에서 가장 잘 알려진 ***BRAF***, ***RAS*** 유전자의 돌연변이를 전체 표본에서 43.75%, 31.25%의 빈도로 발굴 하였으며, ***NFE2L2*** 가 갑상선암 유발 유전자로 제시되었다. ***TERT*** 프로모터 변이는 기존의 보고와 유사하게 75%의 빈도를 보여, 해당 유전자 변이와 공격적인 갑상선암의 관계를 다시 한번 확인 할 수 있었다. 또한, 10 개의 표본에서 발암 유전자 (***EIF1AX***, ***AKT1***, ***PIK3CA***), 종양억제 유전자 (***TP53***, ***BRCA1***, ***PTEN***, ***TET1***, ***LATS1/LATS2***), 염색질조절 유전자 (***ATRX***), 메틸전이효소 유전자 (***KMT2C***), 및 스플라이시오좀 유전자 (***U2AF1***) 등에 존재하는 돌연변이를 발굴하였다. ***EIF1AX*** 는 ***RAS*** 돌연변이가 발견된 표본 5 개 모두에서 발견되었으며, 그중 3 개 표본은 종양억제 유전자인 ***CDKN2A/CDKN2B*** 의 결실을 동반하였다. 또한, 미분화 갑상선암과 함께 분화갑상선암의 전사체 및 DNA 메틸화를 분석을 시행하여 분화갑상선암에서 미분화갑상선암으로 발달 시 발생하는 전사체 및 후성유전체의 변이를 관찰하였다. 암 유발 돌연변이 유전자에 따라 전사체 발현 양상이 ***BRAF***-like, ***RAS***-like, 및 NBNR 로 구분되는

분화갑상선암과 다르게, 미분화갑상선암은 서로 다른 암 유발 돌연변이 유전자를 갖고 있어도 ATC-like 의 형태로 동일한 전사체적 특징을 나타내었으며, CpG island 과메틸화가 동반되는 것이 발견되었다. 또한, 종양면역치료의 대상으로 알려진 IL-6, PD-L1, PD-L2 과 후성유전체를 조절하는 *DNMT1* 과 *DNMT3B* 의 발현이 증가함을 확인하여 미분화갑상선암의 새로운 표적 치료 대상을 제시하였다. 종합하여, 본 연구는 차세대 염기서열 분석 방법을 활용하여 미분화갑상선암과 분화감상선암의 대규모 유전체, 전사체, 및 후성유전체 정보를 통합 분석한 첫 연구이며, 이를 통해 분화갑상선암의 새로운 진단 및 분류 기준과 미분화갑상선암의 발생 기전과 표적 치료 대상이 될 수 있는 신호전달 과정을 규명하였다는 의의가 있다.

* 본 논문의 첫번째 연구는 Plos Genetics 에 출판된 내용임 [1]

주요어: 갑상선암; 미분화 갑상선암; 차세대 염기서열 분석; 전장 유전체; 전사체; 후성유전체

학번: 2014-21325